IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 29, 2021, accepted November 18, 2021, date of publication November 30, 2021, date of current version February 3, 2022.

Digital Object Identifier 10.1109/ACCESS.2021.3131506

Image Processing Based Approach for False Data Injection Attacks Detection in Power Systems

HAMED MOAYYED[1], MOSTAFA MOHAMMADPOURFARD[2], CHARALAMBOS KONSTANTINOU[3], (Senior Member, IEEE), ARASH MORADZADEH[4], (Student Member, IEEE), BEHNAM MOHAMMADI-IVATLOO[4], (Senior Member, IEEE), AND A. PEDRO AGUIAR[1], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, University of Porto, 4099-002 Porto, Portugal
[2]Department of Electrical and Computer Engineering, Sahand University of Technology, Tabriz 5147896, Iran
[3]Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
[4]Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz 15731, Iran

Corresponding author: Mostafa Mohammadpourfard (mohammadpourfard@sut.ac.ir)

ABSTRACT With more sensors being installed by utilities for accurate control of power grids, there is a growing risk of vulnerability to sophisticated data integrity attacks such as false data injection (FDI), circumventing current bad data detection schemes resulting in inaccurate state estimation solutions. While diverse automated detectors to battle FDI have been grown, such methodologies underestimate the strong analytical abilities of humans. This is while most proposed automated methods need observant human control. Although automated methods provide opportunities to improve scalability, humans can cope with exceptions and new attack trends. In this paper, to address the ever-increasing cyber-attack challenge in power systems, a visualization based attack detection framework using deep learning techniques is developed to provide human security researchers with improved techniques to uncover trends, identify outliers, recognize correlations, and communicate their results. Specifically, we first encode multivariate systems state time-series data into 2D colored images and then utilize a carefully designed deep convolutional neural network (CNN) classifier. The proposed method is developed to allow network operators to immediately capture and visually understand the statistical features of a network attack at a glance. The proposed method has been evaluated on the IEEE 14-bus and IEEE 118-bus systems. Our experiments show that the proposed framework accomplishes high classification accuracy.

INDEX TERMS Cyber-attacks, deep learning, image processing, smart grid, false data injection attacks, visualization.

I. INTRODUCTION

The future intelligent grid is extremely dependent on computing, communication, and control technologies to ease the control and operation of the power grid. However, this heavy dependence leaves the modern power grid susceptible to a broad range of cyber-attacks that can decrease the stability of smart grids and ultimately undermine vital national infrastructural sectors, causing significant market failures, civil disruption and considerable financial damage [1]–[4]. Therefore, the demand for safeguarding critical smart grid infrastructures from assorted cyber-attacks is a place of growing anxiety. The cyber-attacks on the intelligent grid can typically be divided into two classes: 1) Physical attacks: In this type of attacks, energy grid parts such as generators and circuit breakers are targeted to change electric power topology that could directly result in power outages and cascade failures [5]–[7]. But, they can still be quickly identified even if the associated protective instruments that record physical elements condition are also jeopardized [8]. 2) Cyber-attacks: This type of attacks, which generally are hard to detect if attack vectors are well-organized, try to delude power system operation by targeting the supervisory control and data acquisition (SCADA) program, which can trigger implicit financial damage and endanger the security of the power system [9]–[12].

The associate editor coordinating the review of this manuscript and approving it for publication was Jiafeng Xie.

A means of organizing such an attack is regarded as false data injection attacks (FDIAs), which presents a significant danger to the credible operation of the smart grids by undermining the integrity of the state estimation results and must be identified and discarded immediately in order to deter severe financial impacts. Various data-driven detectors have been suggested [13] to safeguard the state estimation solutions against FDIA. Generally speaking, the developed automated detectors first approximate the data distribution of past observations and use the calculated distribution to catch potential attacks that result in intense deviation from the reference distribution. For example, in [14], authors have proposed a deep belief network combined with the Gaussian-Bernoulli deep Boltzmann machine to identify FDIA. Margin Setting algorithms were used in [15] to identify FDIA. An ensemble-based classifier using Deep Neural Network (DNN) and Decision Tree (DT) is proposed in [16] to detect FDIA. The work in [17] has discussed the binary classification problem of identifying FDIA utilizing supervised learning algorithms such Support Vector Machine (SVM). Recurrent Neural Networks (RNNs) are used in [18] for binary classification of falsified measurements and normal ones.

While most of the developed automated detectors need meticulous human intervention, no emphasis has been given to the analytical capabilities of the power grid operators. This paper aims to bridge this gap by providing the power grid security analysts with a carefully designed attack detection tool exploiting deep learning techniques to discover patterns and find cyber-anomalies. The proposed method takes advantage of the inherent strengths of a deep learning-based detector and power system operators perceptual to identify legitimate data patterns and rapidly catch and visually apprehend the attacked measurement at a glance. While over the last few years, the problem of detecting cyber threats by visualization has been extensively explored in the traditional information technology systems [19], [20], it has not been widely studied in the context of the power systems yet. Moreover, we tackle the multi-class classification issue where the designed framework can discriminate between different attack classes (attack to different systems states), which makes our solution robust than binary classification solutions that have only two class labels: attack and normal.

Data visualization approaches and image processing techniques play an important role when it comes to data analysis projects. Physically capturing the internal representations of data brings significant advances in pattern discovery and anomaly detection. There are few research efforts to implement data visualization technologies in power systems [21], [22]; however, more attention should be paid to this important goal.

Following this trend, the authors recently presented a novel visualization approach for fault localization in power transformer windings [23]. The presented results confirmed the efficiency of data visualization and the accuracy of this approach.

In this paper, an attack detection framework based on mapping the system states data to 2D images is proposed. The framework utilizes a deep convolutional neural network (CNN), which has obtained significant success in the image processing area. While a significant portion of the currently developed detectors relies on one-dimensional signals, this paper transforms the system state time-series data into 2-dimensional representations and uses the deep CNN classification. Image representation of system states brings up various types of features that are not accessible for one-dimensional state vectors leading to boosting the recognition rate of the detector. Therefore, we treat the attack detection task as a texture image recognition problem.

The rest of the paper is organized as follows: Section II illustrates the problem settings for our work. Section III details the proposed deep learning-based framework and computer vision. Section IV explains the visualization in power systems. The proposed approach and results are presented in Section V. Finally, the paper is concluded in Section VI.

II. POWER SYSTEM STATE ESTIMATION AND FDIA

A. POWER SYSTEM NONLINEAR STATE ESTIMATION

The relation between the measurement vector $\mathbf{z} = [z_1, z_2, \cdots, z_m]^T \in \mathbb{R}^{m\times 1}$ and the state vector $\mathbf{x} = [v_1, \cdots, v_n; \theta_1, \cdots, \theta_n]^T \in \mathbb{R}^{2n\times 1}$ in AC power flow model which contains the voltage magnitudes and phase angles, is given by [24]:

$$\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}, \tag{1}$$

where $\mathbf{h}(\cdot)$ is a nonlinear function relating $\mathbf{z}$ to $\mathbf{x}$. $\mathbf{e} = [e_1.e_2, \cdots, e_m]^T \in \mathbb{R}^{m\times 1}$ is the measurement error vector which follows Gaussian distribution with zero mean. The estimated system state $\widehat{\mathbf{x}}$ is calculated by minimizing the weighted least squares criterion, yielding:

$$\widehat{\mathbf{x}} = \arg\min_{x}[\mathbf{z} - \mathbf{h}(\mathbf{x})]^T R^{-1}[\mathbf{z} - \mathbf{h}(\mathbf{x})], \tag{2}$$

where $\mathbf{R}$ is the measurement error covariance matrix. After the state estimation step, ℓ_2-norm detector is used to identify existence of bad measurement by checking if the following condition holds [24]:

$$\|r\| = \|\mathbf{z} - \mathbf{h}(\widehat{\mathbf{x}})\| \geq \tau, \tag{3}$$

where τ is a predetermined detection threshold.

B. FDIA MODEL OF NONLINEAR STATE ESTIMATOR

In FDIA [9], an intruder, who has a priori knowledge of the network topology and also can simultaneously access and manipulate limited amounts of real-time measurements, can pass the bad data detection (BDD) considered in (3). Let $\mathbf{a} \in \mathbb{R}^{m\times 1}$ be the attack vector. After FDIA, the state estimation will get an erroneous system state $\widehat{\mathbf{x}}_a = \widehat{\mathbf{x}} + \mathbf{c}$ from the manipulated measurement data where $\mathbf{c} \in \mathbb{R}^{n\times 1}$ is the injected arbitrary errors into the true estimates of the system $\widehat{\mathbf{x}}$. More precisely, if the real measurement $\mathbf{z}$, could

circumvent the BDD, the attacked measurement $\mathbf{z}_a = \mathbf{z} + \mathbf{a}$ could also bypass BDD if the following holds:

$$\mathbf{a} = \mathbf{h}(\widehat{\mathbf{x}} + \mathbf{c}) - \mathbf{h}(\widehat{\mathbf{x}}). \tag{4}$$

This because under (4), the measurement residual $\mathbf{z}_a$ is the same as that of original measurement $\mathbf{z}$ since:

$$\begin{aligned} \|r_a\| &= \|\mathbf{z}_a - \mathbf{h}(\widehat{\mathbf{x}}_a)\| = \|\mathbf{z} + \mathbf{a} - \mathbf{h}(\widehat{\mathbf{x}} + \mathbf{c})\| \\ &= \|\mathbf{z} + \mathbf{a} - \mathbf{h}(\widehat{\mathbf{x}} + \mathbf{c}) + \mathbf{h}(\widehat{\mathbf{x}}) - \mathbf{h}(\widehat{\mathbf{x}})\| \\ &= \|\mathbf{r} + \mathbf{a} - \mathbf{h}(\widehat{\mathbf{x}} + \mathbf{c}) + \mathbf{h}(\widehat{\mathbf{x}})\| = \|r\| \leq \tau. \end{aligned} \tag{5}$$

Therefore, from the above, one can conclude that the BDD alone is not suitable to prevents such attacks. In this paper, we provide a method that allow network operators to immediately capture and visually understand the statistical features of a network attack at a glance.

III. DEEP COMPUTER VISION

In recent years, Artificial Intelligence (AI) has been highly proposed to enhance the relationship between humans and machine capabilities, and many aspects of this field are examined. One of many such areas is the domain of Computer Vision. Scholars are utilizing from Deep Learning to advance Computer Vision such algorithm as Convolutional Neural Network (CNN), primarily, which is introduced as a class of deep learning neural networks.

CNNs are inspired by brain's visual cortex and provide spatial data and use an ad-hoc architecture. Many different neurons have a small field of local receptive and are only sensitive to visual stimuli set in a limited area of the visual field in the visual system. CNNs only react to images of horizontal lines and other ones can react to lines with district orientations. In addition, different neurons possess larger receptive fields that react to more complex roles consisted of the lower-level patterns. It was determined that higher-level neurons are based on the outputs of neighboring lower-level neurons. Hence, for any area of the visual field, the mentioned powerful architecture is very impressive in order to detect the whole series of complex patterns.

Deep learning and CNNs had the primary role in obsoleting traditional computer vision techniques, which had been undergoing progressive development. These two methods will be detailed in the following sections, and related attempts in literature will be introduced.

A. DEEP LEARNING

Deep learning algorithms have been introduced as a subset of machine learning algorithms that can be used for finding multiple states of distributed representations. It is commonly used in traditional artificial intelligence domains like computer vision [25], semantic parsing [26], natural language processing [27], etc. For more use of deep learning in current applications, there are three considerable reasons: the dramatically lowered cost of computing hardware and the significant progress and the dramatically enhanced chip processing

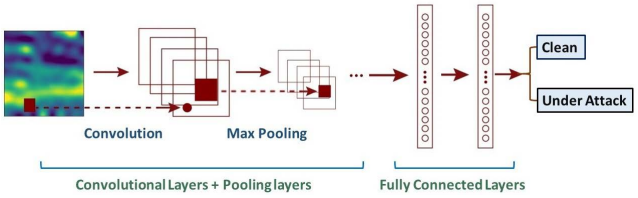

FIGURE 1. The pipeline of the general CNN architecture.

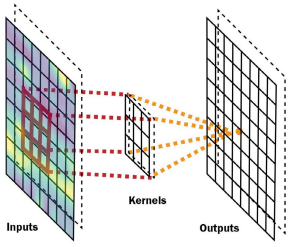

FIGURE 2. The operation of the convolutional layer.

abilities such as CPU or GPU units in the machine learning algorithms [28].

Deep learning advances have been widely reviewed and addressed in recent years. In [29], the challenges in the case of deep learning research investigated and suggested a few forward-looking research directions, whereas [30] focused the dramatic inspirations as well as technical contributions in a historical timeline format. Deep networks can extract proper parameters, whereas jointly accomplishing discrimination so are useful for computer vision [31]. Recently, deep learning approaches, among ImageNet Large Scale Visual Recognition Challenge (ILSVRC) competitions, are very regarded by scientists and provided highly accurate scores [32].

B. CONVOLUTIONAL NEURAL NETWORKS (CNNs)

The Convolutional Neural Networks (CNN) is known as one of the most considerable deep learning methods in which one can train multiple layers in an end-to-end manner [33]. CNN is determined to be very impressive and is considerably used in diverse computer vision usages. Figure 1 shows the pipeline of the whole CNN architecture.

Commonly, as can be seen in Fig. 1, a CNN consisted of a hierarchical neural network in which convolutional layers alternate with pooling layers after many fully connected layers. The function of the three mentioned layers and recent developments applied to those layers are provided in the following.

Convolutional layers; Figure 2 shows that CNN can use many kernels in order to convolve the images along with the intermediate feature maps, which can make different feature maps.

The operation of convolution has basic advantages. In the case of same feature map, a weight sharing mechanism can decrease the number of elements and, among neighboring

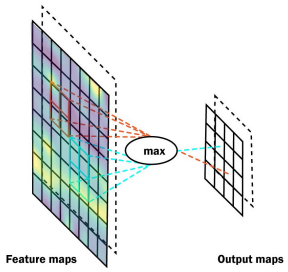

FIGURE 3. The operation of the max pooling layer.

pixels, correlations can be learned by local connectivity as well as invariance to the location of the object.

Pooling layers Particular, a convolutional layer is necessary for a pooling layer. A convolutional layer is utilized for decreasing the dimensions of feature maps along with network elements. Due to the fact that the computations of pooling layers consider neighboring pixels, pooling layers are additionally translation-invariant like convolutional layers. From many different strategies, average pooling and max pooling have been highly utilized. A max pooling process is shown in Fig. 3.

Fully − connected layers; These layers converting the 2D feature maps in a one-dimension feature vector, in the case of more feature representation.

Classification layer; The last layer of the CNN structure is the classification layer, which categorizes the trained data in fully-connected layers based on the labeled targets. Classification in this layer is done by the Softmax function.

In a CNN, fully-connected layers provide feed-forward the neural network in a vector by a pre-defined length. In this regard, one can take fully-connected layers as a feature vector in the case of follow-up processing or feed-forward the vector in specified number categories in the case of image classification.

IV. POWER SYSTEMS VISUALIZATION

For describing a phenomenon, a picture is better than a thousand words because it conveys a lot of detail through a visual representation of information while providing instant access and a meaningful approach. The human brain is more sensitive to visual processing compared with any other sense. In the case of understanding, visual processing is ''broadband'' access. Our mind's ability to process visual input rapidly causes data visualization, an appropriate and commonly effective tool that enables us to change data into knowledge and information.

For producing much data, there are many computational devices in power systems that use many complex algorithms consist of significant information in the case of control to operation. The data that is automatically arrived from the tools used for power system operation in control centers. After that, operators can analyze these outcomes and, according to their interpretation, applied appropriate actions that are commonly under time pressure.

The power system of computational tools make data that consist of many obstacles to the comprehension of human. Operators need mental representations of the data to be created and, after that, analyze them in order to extract more efficient information. A visualization way (i.e., the physical understanding of these internal representations) can make a considerable advancement in the facility of comprehension [10], [22].

Now, for making this possible, there is proper hardware for modern graphical user interfaces, but enough representations of these approaches are not currently implemented. This paper applied them to mapping the system states data to 2D images and implementing it to detect FDIAs.

V. PROPOSED APPROACH AND RESULTS

This section includes: a) how the mapping of system states to 2D images and present a novel technique to detect cyber anomalies, b) the multi-class classification between various attacks to different systems states, and detailed all accuracy results based on CNN approach.

The simulations are implemented on various systems. We use 14-bus and 118-bus test systems mainly to demonstrate the performance. The historical data have been preprocessed by MATPOWER. Real-world load data were obtained from the New York Independent System Operator (NYISO) to improve the accuracy of subsequent simulations and generate time-series data. It is presumed that the intruder can compromise a portion of the sensor readings through the communication device, modifying the voltage state. It is then assumed that the hacker would eventually insert false data into each system state. Note that for 118-bus system, false data are injected into 19 randomly selected buses. For every attacked state, two injection amounts 90% and 110% of the real value are conducted. For example, 90% indicates that the manipulated state variable by the adversary is 10% less than the actual value.

A. VISUALIZATION METHODOLOGY

The 2D images of system states were formed by using Matplotlib, which is a modern library in Python [34]. This library is a 2D graphics environment that could be applied for different application development to generate 2D images across user interfaces and operating systems. The detailed description of the visualisation method used is as follows.

The visualization technique applied can be composed in three main steps as follows: 1- According to the order of number data, all data are mapped to a square block including mini-squares. 2- By applying Matplotlib, a cross-platform data visualization library in Python, each number is assigned a specific color based on the amount of that value in the mini-squares. 3- The output 2D color image is used as input to the CNN network for the image processing approach. Figure 4 represents this process of converting data into an image for a simple case of a vector between 1 and 4. As shown, the values of this vector are mapped to mini squares, and then each number is assigned a specific color. The lowest value

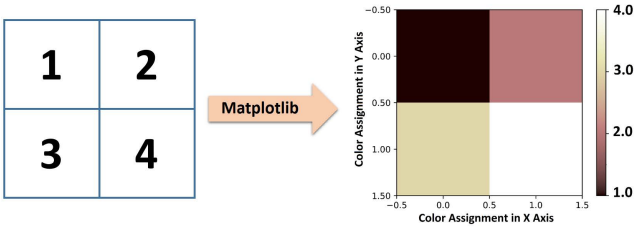

FIGURE 4. Process of visualizing data into 2D images.

has a lighter color, while 1 is mapped to white, and the highest value of 4 is assigned the darkest color, black.

B. MODEL ARCHITECTURE

The network architecture was based on the CNN, as shown in Fig. 1. The network was designed to input the 2D image from the Matplotlib library and output the multi-label image classifications. The network consists of 10 layers including 5 convolutional layers (Conv). Conv1 and Conv2 are used to tune the convolutional features, so the number of feature maps of both layers is the same, with a kernel size of 32 and equipped with activation functions of Tanh. Conv3 and Conv4 play the role of tuning the convolutional features according to the feature module, therefore the parameter settings are the same and the kernel size is set to 64 and equipped with the activation function of Tanh. Finally, Conv5 plays the important role of reconstructing feature maps into channel output. Thereby, the kernel size of Conv5 is set to 128. Moreover, in order to overcome the overfitting problem, the last convolutional layer (Conv5) were equipped with a ReLU activation layer and a dropout (0,25) layer. The CNN were designed using the Tensorflow deep learning framework and trained by the Adam optimization algorithm.

C. CLASSIFICATION METRICS

The performance of the Deep CNN technique is evaluated using F1-score and accuracy indicators [35]. The F1-score is one of the statistical evaluation indicators which is utilized for performance evaluation in classification applications of machine learning techniques and is calculated as follows:

$$\mathbf{F1} - Score = \frac{\mathbf{2} \times \mathbf{p} \times \mathbf{r}}{\mathbf{p} + \mathbf{r}}, \tag{6}$$

where $\mathbf{p}$ and $\mathbf{r}$ represents the precision and recall, respectively, and are obtained as following equations:

$$\mathbf{p} = \frac{\mathbf{t}p}{\mathbf{t}p + \mathbf{f}p}, \tag{7}$$

$$\mathbf{r} = \frac{\mathbf{t}p}{\mathbf{t}p + \mathbf{f}n}, \tag{8}$$

where $\mathbf{t}p$ is the true positive and $\mathbf{t}n$ demonstrate the true negative. $\mathbf{f}p$ and $\mathbf{f}n$ depicts the false positive and false negative, respectively. The accuracy indicator can be utilized with different calculations to evaluate the performance of both regression and classification algorithms. In this paper, this

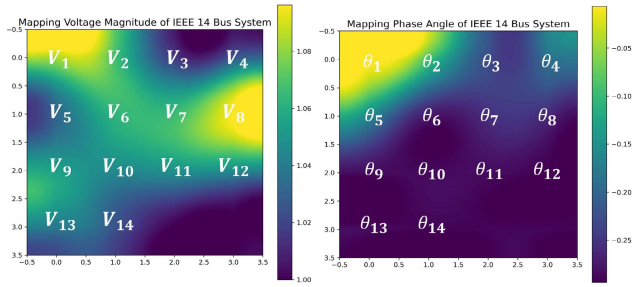

FIGURE 5. Mapping of IEEE 14-bus system states data to 2D images. Voltage magnitude and phase angle map to a square 2D according to an assigned distinct location.

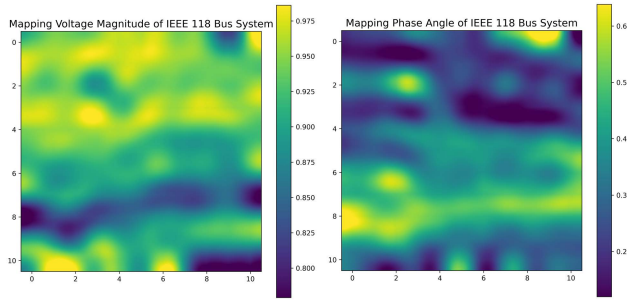

FIGURE 6. IEEE 118-bus system state variables representation within 2D images. Each state variable is visualised in a specific area of the image.

indicator is employed for classification applications according to the following equation:

$$\mathbf{A}ccuracy = \frac{\mathbf{t}p + \mathbf{t}n}{\mathbf{t}p + \mathbf{f}p + \mathbf{t}n + \mathbf{f}n}. \tag{9}$$

D. VISUALISATION RESULTS

For mapping the state variables and generating a 2D image, the system states were separated into voltage magnitude and phase angle and project to a square 2D image. As depicted in Fig. 5 for the IEEE 14- bus, each 2D image depicts the colormap assigned to each state variable. The same approach was applied to display the states variables of IEEE 118-bus system, as shown in Fig. 6. The voltage magnitude and phase angle are mapped to separate images and assigned to a color corresponding to their value.

When FDIA occurs, an attacker compromises measurements from the grid sensors in such a way that undetected errors are introduced into estimates of state variables such as voltage magnitude and phase angles. For each attack, one system state variable is decreased or increased by a certain percentage of its original value.

As a consequence of FDIA, the anomaly caused in system states would change the colormap images formed by assigned state variables. Identifying such pixels with anomalous colors can aid in both detection of the FDIA and even important knowledge that may lead to the target of this detection mission.

Identifying anomalies would be difficult for the human eye system. The proposed method relies on detecting anomalies

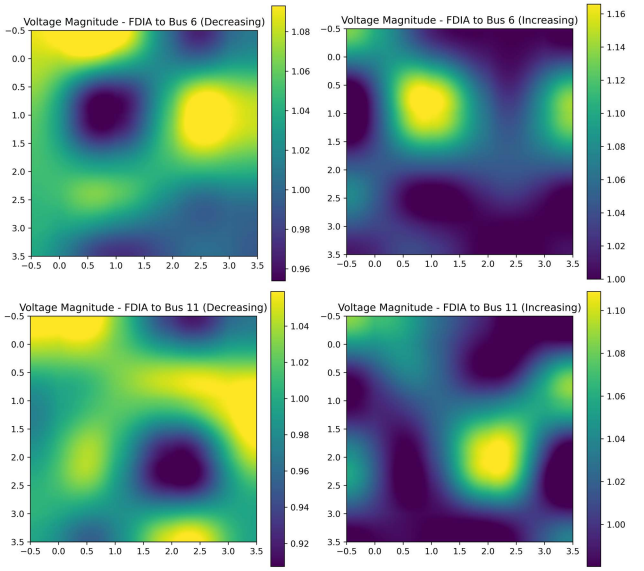

FIGURE 7. Visualization of the IEEE 14-bus system's voltage magnitude, when decreasing and increasing FDIA happen to bus 6 and 11.

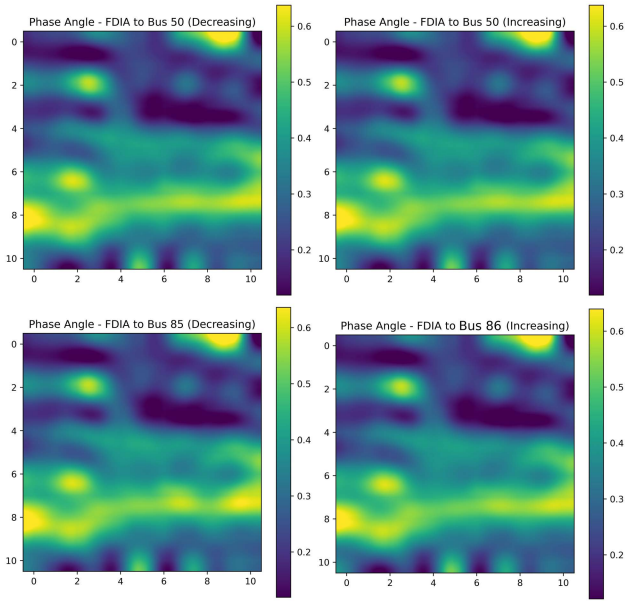

FIGURE 8. Visualization of the IEEE 118-bus system's phase angle, when decreasing and increasing FDIA happen to bus 50 and 86.

concerning the residual of corresponding images. Therefore, from the input image, a residual image containing everything that does not repeat would be extracted.

Figure 7 shows the obtained images of the IEEE 14-bus system's voltage magnitude in the presence of FDIA in buses 6 and 11, respectively, for a decreased and an increased value of the real one, respectively. By comparing these images with the normal system images shown in Fig. 5, it is possible to check the anomalies that appear in new images. In particular, it can be seen that there exist regions in the image that are not conforming with primary images, and this irregularity is different for FDIA to a particular bus and even the kind of attack as decreasing or increasing.

On larger systems, quickly finding these anomalies in images would be more difficult. Figure 8 represents images for the IEEE 118-bus system's phase angle when decreasing and increasing FDIA happen to bus 50 and 86. An in-depth focus, and comparing these images with normal system images, confirm the difference in color arrangement when FDIAs happens.

Anomaly detection of the images would lead to extract useful knowledge about the FDIA. The detection methodology that has been proposed was depicting the residual images (extracted from normal system image) in which anomalies prevail.

Figure 9 shows the outcomes of the anomaly detection approach applied to the images in Fig. 7. Looking into residual images gives an accurate perspective about the FDIA. Firstly, the presence of an abnormality in the colormap is a worth symptom of FDIA. Secondly, the position of this irregularity could determine the exact area of the system, which is under attack. Last but not least, the dominant color of the residual image can identify the decreasing or increasing of the attack.

As shown in Fig. 9, in IEEE 14-bus system, the anomaly detection approach leads to the detection of FDIA, address the accurate position of the system under attack (bus 6 and bus 11) and even demonstrate the decreasing and increasing of the attack. On a larger case, IEEE 118-bus system, detection of abnormalities method applied to the images in Fig. 8. As depicted in Fig. 10, the residual images give essential knowledge about the FDIA, including detecting the position of attack and identifying which kind of attack was occurred.

The visualisation approach was also compared with Principal component analysis (PCA). Dimensionality reduction and visualization using PCA is a well-known method that has been applied to many tasks. The performance of the proposed method based on CNN was compared with PCA in two different scenarios, FDIA to Bus 15 (Increasing) and FDIA to Bus 50 (Decreasing). As shown in Fig. 11, the generated images appear from the CNN model's output can clearly demonstrate the area under attack. However, the visualization of the system state data with the same attacks is also shown in Fig. 11. The axes of these graphs are principal components (PCs) calculated after applying PCA on the state vectors. As one can see, the standard operation data and the manipulated data are interwoven.

E. CLASSIFICATION RESULTS

The use of deep learning techniques for classification and regression applications requires an input dataset. In this paper, voltage and phase angle data obtained from the 14-bus IEEE and 118-bus IEEE test systems related to the healthy state and FDIA are available and examined. To use the

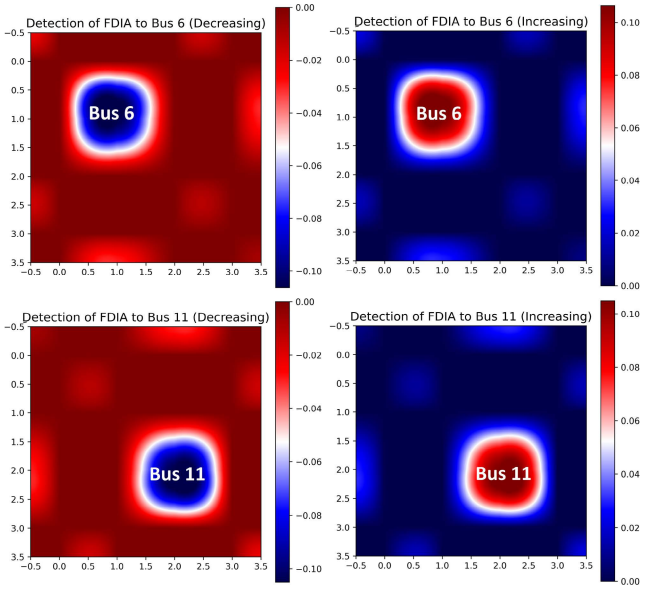

FIGURE 9. Detection of FDIA in IEEE 14-bus system. The residual of normal system's with the system under FDIA leads to discover the attack, the area under attack, and even distinguish decreasing or increasing FDIA.

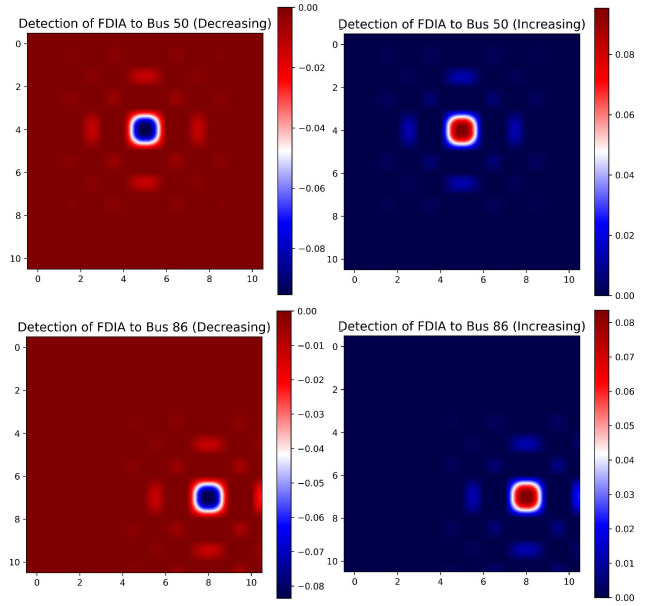

FIGURE 10. Detection of FDIA in IEEE 118-bus system. Displaying the difference image between the normal system and the system under attack demonstrates the desired knowledge about FDIA.

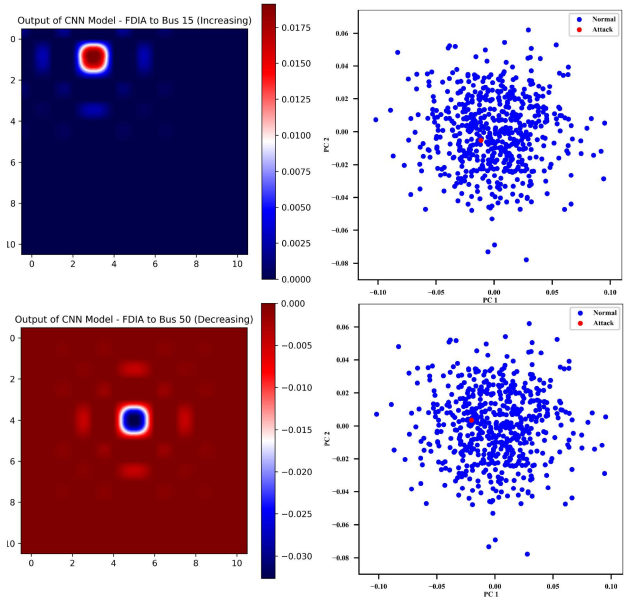

FIGURE 11. Comparing the output of CNN model and PCA in two different cases, FDIA to Bus 15 (Increasing) and FDIA to Bus 50 (Decreasing).

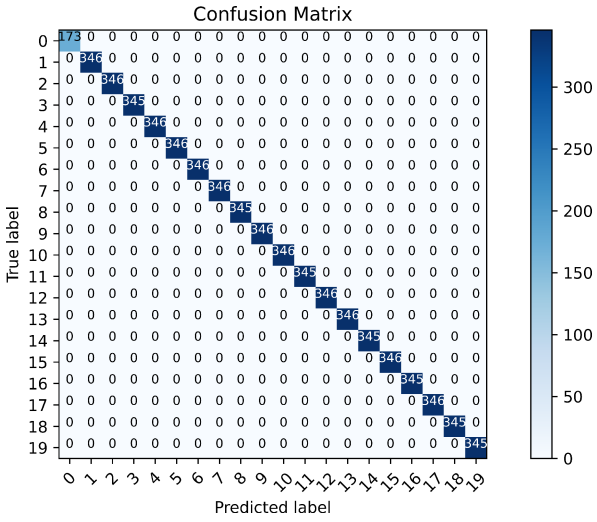

FIGURE 12. Confusion Matrix of classifying the FDIA corresponding to the bus 118 of IEEE 118-bus in voltage data.

proposed method for classification and locate each FDIA, the mapped graphs of the voltage values and phase angles formed the input dataset. Each of these data was employed separately to classify and identify the FDIA.

For detecting the attack in each of the IEEE test systems using voltage-related values, the CNN was trained with 70% of the data and tested with the rest of the data. In the identification of the FDIA by phase data, 80% of the data was selected for training and 20% of the data was considered for test the network.

After training and test the networks, the results of attack detection and classification are presented in the forms of confusion matrices. The results of identifying and classifying the FDIA corresponding to the 14-bus test system using the voltage data done with 100% accuracy and F1-score of 100%. This evaluation for phase data was 93.34% and 93.00% for accuracy and F1-score, respectively.

On larger systems, IEEE 118-bus system, the results of the FIDA diagnosing and classifying using the voltage and phase angle data are presented in Figures 12 and 13, respectively.

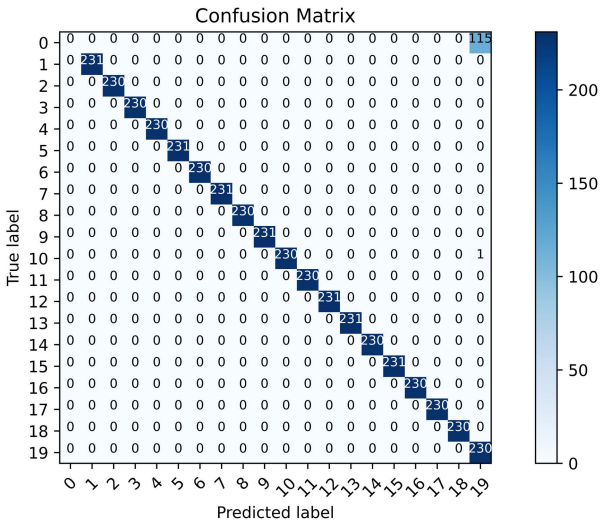

FIGURE 13. Confusion Matrix of classifying the FDIA corresponding to the bus 118 of IEEE 118-bus in phase data.

According to the presented results in the above figures, it can be seen that networks had a better performance in detecting FDIA based on voltage data and were able to classify attacks with 100% accuracy. In the case of phase data, attacks were identified and classified with an accuracy of 97.44% for the 118-bus test system. From the detection results, it can be seen that considered attacks in power systems have more effects on system voltage and the proposed method was able to extract a good pattern of behavior from voltage data in both test systems. The results presented for the phase angle data have also been performed with acceptable performance, but the results confirm the existence of salient features in the voltage data and the high ability of the proposed method to extract them.

All training and testing procedures, including image generation, were performed on a laptop with a COREi7-Inetel processor and 16.0 GB installed RAM. The algorithm was run in PyCharm IDE, Edition 2020.1.2, a popular Python development platform. The CPU runtime for the proposed detection algorithm was reported to be 32 seconds, which shows the superiority of the algorithm in terms of computational time in addition to its efficiency and robustness.

By using the proposed method, the generated images can be analyzed by the system operator in the control center. The anomalies that appear in the images could be easily detected by eye and used as a tool that can be integrated into the control center as an alarm system. Instead of comparing time series and performing complex calculations, the attack could be identified simply by looking at the 2D image using our proposed method, as shown in Figures 9 and 10.

In this paper, a comparative approach for detecting and classifying attacks is presented to express the high efficiency and performance of the proposed procedure. It should be noted that in machine learning and deep learning applications, comparisons should be made with caution and for similar

TABLE 1. Comparison of the Deep-CNN and SVM methods.

		14-bus test system		118-bus test system	
Technique	Indicator	Voltage	Phase	Voltage	Phase
Deep-CNN	Accuracy(%)	100	93.34	100	97.44
	F1-Score(%)	100	93.00	100	97.00
SVM	Accuracy(%)	97.23	92.36	97.02	95.17
	F1-Score(%)	97.00	92.00	97.00	95.00

data. Accordingly, in this paper, one of the most widely used machine learning techniques called SVM, which is specifically utilized to solve classification problems, is applied to voltage and phase angle data related to test systems. As mentioned, the comparison of the Deep-CNN and SVM methods is done by the same data to evaluate the results with high accuracy. Table 1 makes this comparison based on the accuracy and F1-score indicators for both methods.

VI. CONCLUSION

While visualization for cyber-security is trying to fix the critical challenges of information security by enabling people through data visualization, it has not received enough attention in the smart grid cyber-security context yet. This paper proposed a deep learning-based visualization technique to detect injected false data into power system measurements, which lets the grid operator to uncover data trends and obtain useful insights over legal data patterns to spot suspicious patterns when FDIA occurs. In particular, the system state signals are first converted into 2D images and then processed by a carefully designed deep CNN framework. The proposed method enables the grid operators to visualize multi-dimensional power system measurements using 2D images, and also can dynamically capture various aspects of time-series characteristics and classification simultaneously. Test results indicate that the proposed method can reliably detect and localize most of the FDIA over different power system networks. Experimental results also show the superiority of the proposed method over current FDIA detectors such SVM and also conventional visualization-based attack detectors such as PCA.

REFERENCES

[1] B. Li, G. Xiao, R. Lu, R. Deng, and H. Bao, "On feasibility and limitations of detecting false data injection attacks on power grid state estimation using D-FACTS devices," *IEEE Trans. Ind. Informat.*, vol. 16, no. 2, pp. 854–864, Feb. 2020.

[2] M. Mohammadpourfard, A. Khalili, I. Genc, and C. Konstantinou, "Cyber-resilient smart cities: Detection of malicious attacks in smart grids," *Sustain. Cities Soc.*, vol. 75, Dec. 2021, Art. no. 103116.

[3] M. Mohammadpourfard, I. Genc, S. Lakshminarayana, and C. Konstantinou, "Attack detection and localization in smart grid with image-based deep learning," Oct. 2021, *arXiv:2110.11007*.

[4] I. Zografopoulos, J. Ospina, X. Liu, and C. Konstantinou, "Cyber-physical energy systems security: Threat modeling, risk assessment, resources, metrics, and case studies," *IEEE Access*, vol. 9, pp. 29775–29818, 2021.

[5] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2218–2234, May 2019.

[6] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "A framework for cyber-topology attacks: Line-switching and new attack scenarios," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1704–1712, Mar. 2017.
[7] G. Chen, Z. Y. Dong, D. J. Hill, and Y. S. Xue, "Exploring reliable strategies for defending power systems against targeted attacks," *IEEE Trans. Power Syst.*, vol. 26, no. 3, pp. 1000–1009, Aug. 2011.
[8] R. Deng, P. Zhuang, and H. Liang, "CCPA: Coordinated cyber-physical attacks and countermeasures in smart grid," *IEEE Trans. Smart Grid.*, vol. 8, no. 5, pp. 2420–2430, Sep. 2017.
[9] M. Mohammadpourfard, Y. Weng, M. Pechenizkiy, M. Tajdinian, and B. Mohammadi-Ivatloo, "Ensuring cybersecurity of smart grid against data integrity attacks under concept drift," *Int. J. Electr. Power Energy Syst.*, vol. 119, Jul. 2020, Art. no. 105947.
[10] S. Bi and Y. J. Zhang, "Graphical methods for defense against false-data injection attacks on power system state estimation," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1216–1227, May 2014.
[11] G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, "A review of false data injection attacks against modern power systems," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1630–1638, Jul. 2016.
[12] R. Deng, P. Zhuang, and H. Liang, "False data injection attacks against state estimation in power distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2871–2881, May 2018.
[13] X. Wang, X. Luo, Y. Zhang, and X. Guan, "Detection and isolation of false data injection attacks in smart grids via nonlinear interval observer," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6498–6512, Aug. 2019.
[14] Y. He, G. J. Mendis, and J. Wei, "Real-time detection of false data injection attacks in smart grid: A deep learning-based intelligent mechanism," *IEEE Trans. Smart Grid.*, vol. 8, no. 5, pp. 2505–2516, Sep. 2017.
[15] Y. Wang, M. M. Amin, J. Fu, and H. B. Moussa, "A novel data analytical approach for false data injection cyber-physical attack mitigation in smart grids," *IEEE Access*, vol. 5, pp. 26022–26033, 2017, doi: 10.1109/ACCESS.2017.2769099.
[16] A. Al-Abassi, H. Karimipour, A. Dehghantanha, and R. M. Parizi, "An ensemble deep learning-based cyber-attack detection in industrial control system," *IEEE Access*, vol. 8, pp. 83965–83973, 2020.
[17] M. Mohammadpourfard, Y. Weng, and M. Tajdinian, "Benchmark of machine learning algorithms on capturing future distribution network anomalies," *IET Gener., Transmiss. Distrib.*, vol. 13, no. 8, pp. 1441–1455, Apr. 2019.
[18] A. Ayad, H. E. Z. Farag, A. Youssef, and E. F. El-Saadany, "Detection of false data injection attacks in smart grids using recurrent neural networks," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid Technol. Conf. (ISGT)*, Washington, DC, USA, Feb. 2018, pp. 1–5.
[19] M. Celenk, T. Conley, J. Willis, and J. Graham, "Predictive network anomaly detection and visualization," *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 2, pp. 288–299, Jun. 2010.
[20] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with visualization strategy," *Expert Syst. Appl.*, vol. 41, no. 9, pp. 4139–4147, Jul. 2014.
[21] P. Cuffe and A. Keane, "Visualizing the electrical structure of power systems," *IEEE Syst. J.*, vol. 11, no. 3, pp. 1810–1821, Sep. 2017, doi: 10.1109/JSYST.2015.2427994.
[22] V. Miranda, P. A. Cardoso, R. J. Bessa, and I. Decker, "Through the looking glass: Seeing events in power systems dynamics," *Int. J. Electr. Power Energy Syst.*, vol. 106, pp. 411–419, Mar. 2019, doi: 10.1016/j.ijepes.2018.10.024.
[23] A. Moradzadeh, H. Moayyed, B. Mohammadi-Ivatloo, G. B. Gharehpetian, and A. P. Aguiar, "Turn-to-turn short circuit fault localization in transformer winding via image processing and deep learning method," *IEEE Trans. Ind. Informat.*, early access, Aug. 18, 2021, doi: 10.1109/TII.2021.3105932.
[24] A. Abur and A. G. Exposito, *Power System State Estimation: Theory and Implementation*. New York, NY, USA: Marcel Dekker, 2004.
[25] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3642–3649.
[26] A. Bordes, X. Glorot, J. Weston, and Y. Bengio, "Joint learning of words and meaning representations for open-text semantic parsing," in *Proc. Artif. Intell. Statist.*, 2012, pp. 127–135.
[27] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.
[28] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Trans. Signal Inf. Process.*, vol. 3, pp. 1–29, Jan. 2014.
[29] Y. Bengio, "Deep learning of representations: Looking forward," in *Proc. Int. Conf. Stat. Lang. Speech Process.*, 2013, pp. 1–37.
[30] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
[31] Y. LeCun, "Learning invariant feature hierarchies," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 496–505.
[32] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M, Bernstein, and A. C. Berg, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.
[33] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[34] J. D. Hunter, "Matplotlib: A 2D graphics environment," *Comput. Sci. Eng.*, vol. 9, no. 3, pp. 90–95, May/Jun. 2007.
[35] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retr.*, 2005, pp. 345–359.