Power System Event Identification with Transfer Learning Using Large-scale Real-world Synchrophasor Data in the United States

Jie Shi*†, *Member, IEEE,* Koji Yamashita*, *Member, IEEE,* and Nanpeng Yu*, *Senior Member, IEEE*
*Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521
†Department of Systems Engineering, Cornell University, Ithaca, NY 14850
Email: jshi005@ucr.edu, kyamashi@ucr.edu, nyu@ece.ucr.edu

Abstract—The lack of sufficient labeled events and long training time limit the applicability of deep neural network-based power system event identification using synchrophasor data. In this paper, we propose to leverage transfer learning technique to boost the reliability and reduce the required training time of neural classifier for power system event identification. We use the weights of a neural classifier trained on one transmission system as the initial parameters of another neural classifier for a different transmission system. Numerical tests with real-world synchrophasor data from the Eastern and Western Interconnections of the United States show that the proposed transfer learning approach is very effective in not only improving the training reliability but also reducing the training time.

Index Terms—Event identification, transfer learning, phasor measurement unit, deep neural network, graph signal processing.

I. Introduction

Synchrophasor data recorded by phasor measurement units (PMUs) make it possible to detect and classify abnormal power system events in a timely manner. Although deep neural network-based power system detection and classification algorithms have achieved high accuracy, they often require a large amount of power system event training labels and long training time. If a transmission grid operator only has limited event labels and/or historical PMU data, then it is difficult to reliably train a deep neural network that achieves high event classification accuracy. Even if abundant event data are available, training a deep neural network with random initial parameters from scratch could be time and resource consuming. To promote the adoption of deep neural network-based power system event identification solutions, it is crucial to address these two issues in the training process.

In this work, we attempt to mitigate the issues associated with limited event labels and reduce the training time of

deep neural network-based power system event identification algorithms with transfer learning techniques. The key idea of transfer learning is to facilitate the training process of a new model by exploiting the information from a previously trained model of a related task. Equipped with two years of synchrophasor data from hundreds of PMUs across the United States, we investigate the effectiveness of transfer learning by considering the event identification of the Eastern Interconnection and the Western Interconnection as two individual tasks. We would like to answer the question of "Does the neural network trained to identify power system events for the Eastern Interconnection could provide useful information and guidance when building the event identification engine for the Western Interconnection?" The answer is a resounding yes based on the numerical study results with real-world PMU data.

Many power system event detection and identification algorithms based on PMU data have been developed. The related works can be divided into two groups. The first group only detects abnormal events without identifying the event type. The second group performs both event detection and classification. The technical methods used in the first group often falls into one of the following five domains: spectral analysis [1], [2], [3], [4], estimation error [5], [6], spatial correlation variation [7], [8], [9], low-rank property of PMU data [10], [11], and data mining [12]. Event identification using PMU data is a classification problem, which requires a large amount of training data. Majority of the works in the second group do not have access to sufficient labeled real-world PMU data, thereby limiting their effectiveness. For example, the datasets used by [13] and [14] only contain 32 and 57 labeled events, respectively. Case studies in [15] only cover 4 PMUs. Reference [16] focuses only on identifying frequency events.

Our previous work [17] is one of the first studies that have leveraged a large labeled real-world PMU dataset. We proposed a deep convolutional neural network (CNN) based approach to automatically identify power system events in real-time. The deep CNN is trained on two years PMU data, which contain over one thousand labeled power system events from the Eastern Interconnection in the U.S. In this work, we extend [17] by investigating the effectiveness of transfer learning in power system event identification based on deep CNNs. Specifically, we aim to improve the reliability and
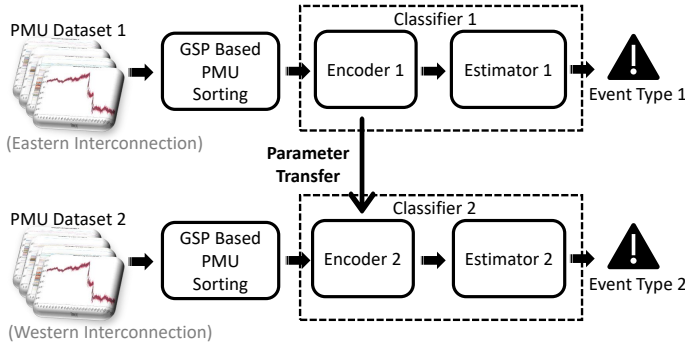
Fig. 1: Overall framework of power system event detection with transfer learning.

computation efficiency of training a deep CNN on the PMU data gathered from the Western Interconnection by exploiting the previously learned model for the Eastern Interconnection.

The contributions of this paper are summarized as follows:

- We validate the effectiveness of transfer learning in the field of power system event identification with large quantity of real-world PMU data.
- We show that the reliability of neural network training for the Western Interconnection is significantly improved by transferring the learned parameters from a model trained on the PMU data from the Eastern Interconnection.
- We demonstrate that transfer learning can notably reduce training time for deep neural networks designed for power system event identification.

The rest of the paper is organized as follows: Section II presents the overall framework and the technical details utilized in this study. Section III validates the proposed transfer learning technique for power system event identification with a large real-world synchrophasor dataset. The conclusions are stated in Section IV.

II. METHODOLOGY

The power system event identification problem with synchrophasor data is treated as a supervised machine learning problem. We propose to enhance and accelerate the training of deep neural network-based event identification model with the transfer learning technique. In the first subsection, the overall framework of the proposed event identification approach with transfer learning is presented. In the second subsection, we provide the details for the key technical methods of the proposed framework.

A. Overall Framework

The overall framework of the proposed power system event identification approach with transfer learning is illustrated in Figure 1. Suppose we have two electric power systems. The top system has a large dataset of synchropahsor data and event labels, whereas the bottom one only gathers a dataset with limited number of event labels. We will first train a CNN-based classifier for the top system and then transfer the learned

parameters of the CNN to the that of the bottom system as initial values. This way, we could learn a decent power system event classifier for the bottom system with a small dataset.

The input features, data preprocessing technique and the power system event classifier are introduced below. The input features include: active power P, reactive power Q, voltage magnitude $|V|$, and frequency f, which are derived based on the voltage and current phasor readings from the PMUs. Measurement matrices can be formed by putting different PMUs' time series data together for each input feature. A 3-dimensional tensor can be constructed by stacking the measurement matrices of P, Q, $|V|$, and f for all power system event and non-event samples. We call these tensors the $PQ|V|f$ tensors. PMU Dataset 1 and PMU Dataset 2 shown in Fig. 1 contain the $PQ|V|f$ tensors and the corresponding event labels.

Instead of directly feeding the 3-dimensional tensor data into the deep neural network-based event classifier, we leverage graph signal processing (GSP) technique to sort the PMUs. To goal of this sorting is to place highly correlated PMUs close to each other, which facilitates the classifier to learn the spatio-temporal correlations among the PMU measurement time series.

The classifier employed in this study is a deep convolutional neural network, which can be considered as a combination of an encoder and an estimator. The encoder strives to transform the input features into meaningful low-dimensional representations, based on which the estimator can interpret and classify the input data accurately. Note that once the training process for the top classifier is finished, we only transfer the parameters of the encoders instead of the whole classifier from one model to the other as shown in Fig. 1.

B. Technical Methods

In this subsection, we provide the technical details of the methods used in the proposed power system event identification approach with transfer learning. Specifically, we present the GSP based PMU sorting technique, the design of the classifier, and the implementation of transfer learning.

1) GSP-based PMU Sorting: The GSP-based PMU sorting technique aims to improve the effectiveness of the parameter sharing scheme of CNN-based event classifier. It strategically rearrange the sequence of PMUs in the given $PQ|V|f$ tensors such that the highly correlated PMU measurement time series are placed close to each other [17].

We only present the final sorting algorithm in this manuscript and refer the interested readers to [17] for the detailed problem formulation and mathematical derivation:

- Step 1: Calculate the Pearson correlation coefficients between the PMUs in the given system;
- Step 2: Build the corresponding graph Laplacian matrix;
- Step 3: Obtain the eigenvalues and eigenvectors of the graph Laplacian by performing eigendecomposition;
- Step 4: Sort PMUs according to the eigenvector that corresponds to the second smallest eigenvalue of the graph Laplacian matrix.
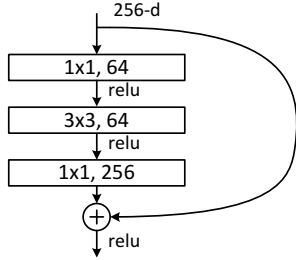
Fig. 2: A typical building block of ResNet50 [18].

Fig. 3: Illustration of a convolutional layer.

2) CNN-based Event Classifier: The classifier is an essential component in the proposed power system event identification framework. It takes in the preprocessed streaming PMU data and outputs the event type estimation. In this work, a deep convolutional neural network called ResNet-50 [18] is adopted as the classifier.

The ResNet-50 can be considered as concatenation of an encoder and an estimator. The encoder covers all the layers except for the last fully connected layer which is left for the estimator. Specifically, the encoder is formulated primarily by stacking individual building blocks. As displayed in Fig. 2, a typical building block is made up of a sequence of convolutional filters. For example, "$3 \times 3, 64$" represents a layer of 64 convolutional filters with dimension of 3×3. Mathematically, a building block can be represented by the following equation:

$$Y_i = f(U_i, \boldsymbol{\theta}_i) + U_i, \tag{1}$$

where U_i and Y_i are the input and output of the ith building block. $\boldsymbol{\theta}_i$ denotes the parameter vector of the ith building block. The parameterized nonlinear function $f(\cdot)$ is the residual mapping to be learned through neural network training. The complete structure of the encoder is built by concatenating an input convolutional layer, a max pooling layer, a series of different building blocks, and a global average pooling layer. We refer the interested readers to [18] for further details.

The estimator is represented by the last layer of the ResNet-50, which is a fully connected layer with outputs normalized by the softmax function. This estimator is intrinsically a linear model, requiring the learned representations to be linearly separable. In other words, the estimator's performance is highly dependent on the quality of encoded representations.

The entire classifier is trained through stochastic gradient descent with Adam optimizer [19]. The training objective is to minimize the categorical cross-entropy loss function:

$$loss = -\sum_{i=1}^{N_c} y_i \cdot \log \hat{y}_i, \tag{2}$$

where N_c is the dimension of output layer, i.e., the number of event classes. y_i and $\hat{y}_i$ are the true value and estimated value of the ith entry in the output array.

3) Transfer Learning: The core idea of transfer learning is to reuse the information of a previously trained model to train
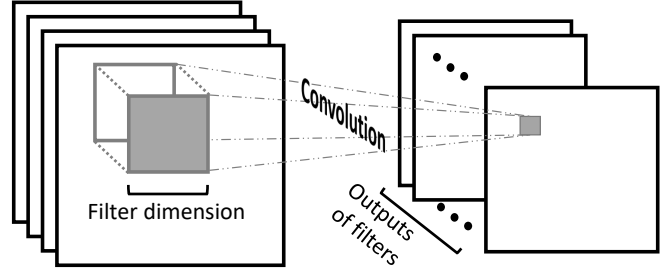
a new model. Rigorously speaking, given two domains D_1 and D_2 as well as two corresponding learning tasks T_1 and T_2, the objective of transfer learning is to improve the learning of the predictive function $f_2(\cdot)$ in D_2 using the previously gained information in D_1 and T_1, where $D_1 \neq D_2$ or $T_1 \neq T_2$ [20].

In this work, domain D_1 and D_2 correspond to the PMU data collected from the Eastern Interconnection and the Western Interconnection of the United States. The learning tasks T_1 and T_2 are to identify the power system events based on the corresponding steaming PMU data. Let $f_1(\cdot)$ and $f_2(\cdot)$ represent the Encoder 1 and Encoder 2 in Fig. 1, respectively. Then the transfer learning technique will facilitate the training process of $f_2(\cdot)$ in terms of speed and robustness by exploiting the previously learned parameters of $f_1(\cdot)$.

Specifically, we first train the Classifier 1 (shown in Fig. 1) with the $PQ|V|f$ tensors calculated from the PMU data of the Eastern Interconnection. After the training session is completed, the learned parameters of encoder $f_1(\cdot)$ are utilized as the initial values of the parameters of encoder $f_2(\cdot)$. The parameters of the Estimator 2 are initialized randomly. Lastly, we train the entire Classifier 2 with the $PQ|V|f$ tensors derived from the PMU data of the Western Interconnection.

It is worth noting that the input features from D_1 and D_2 could have different dimensions due to the different numbers of PMUs. The proposed transfer learning can still work with this dimensionality mismatch since the encoders are built upon the convolutional layers. Fig. 3 illustrates a typical convolutional layer in a CNN. The convolutional filters scan the input features and output the convolution results arranged as stacked matrices. Therefore, the encoders can work with any input dimension with proper zero padding.

III. NUMERICAL STUDIES

In this section, we carry out several numerical studies to quantify the effectiveness of the proposed transfer learning framework for power system event identification. We start this section by providing a brief description of the PMU data used in this work. Then, we evaluate the performance of the GSP-based PMU sorting technique. Finally, we present the settings and results of the proposed transfer learning technique.

A. Data Source

The synchrophasor dataset used in this study includes the measurements from more than two hundred PMUs deployed

across the continental United States. The whole dataset comes from two transmission networks: the Western Interconnection (41 valid PMUs) and the Eastern Interconnection (179 valid PMUs)[1]. The raw measurements, which include positive sequence voltage and current phasors as well as the frequency, are gathered by electric utility companies and regional system operators (RTOs) and then compiled by the Pacific Northwest National Laboratory. Note that the specific PMU locations are not made available to us due to confidentiality concerns.

The raw PMU data are converted to active power P, reactive power Q, voltage magnitude $|V|$, and frequency f, which are put into $PQ|V|f$ tensors. The corresponding power system event labels are created by domain experts from electric utility companies and RTOs. In this work, a total number of 1,147 (1,204) labeled data samples, which include 825 (625) line events, 84 (333) generator events, and 118 (147) oscillation events, and 120 (99) non-events, are provided for the Eastern (Western) Interconnection. The time length of each data sample is 20 seconds with the labeled line or generator event starting time placed in the middle of the window[2]. The reporting frequency of the PMUs is 30 Hz. Therefore, each $PQ|V|f$ tensor in the Eastern (Western) Interconnection has a dimensionality of $[600, 179, 4]$ ($[600, 41, 4]$), where the first entry represents the number of time steps. The second entry corresponds to the number of PMUs. The last entry stands for the four measurement channels: P, Q, $|V|$, and f.

Bad readings and missing values are prevalent in the raw dataset. We follow the same procedure listed in Section IV.B of [17] to detect and replace the bad and missing values. Two issues still exist after the data preprocessing. First, the $PQ|V|f$ tensors are severely imbalanced because of the large number of line events. Second, the starting time stamps of line and generator events are always at the center of the event window, yielding a biased distribution of event timing. To address these two issues, we augment the original $PQ|V|f$ tensors following the same approach discussed in Section IV.D of [17], which produces more balanced datasets for both the Eastern and the Western Interconnections as shown in Table I.

We divide the input $PQ|V|f$ tensors randomly into a training dataset and a testing dataset, which account for 80% and 20% of the total samples, respectively. Note that the training and testing datasets are determined before the data augmentation to avoid potential data leakage. Throughout this study, the training batch size is 16. The total number of training epochs is 200. The learning rate of Adam optimizer is 0.001.

B. Performance of GSP-based PMU Sorting Technique

Fig. 4 shows the weight matrices of the original and the sorted PMU sequence for the 41 valid PMUs in the Western Interconnection. The elements in the weight matrices quantify the correlation levels between corresponding PMUs. It can

---

[1]11 PMUs from the original dataset are considered invalid due to prevalent bad measurements.

[2]Oscillation event samples do not have labeled starting times in dataset. The oscillation signature covers the entire 20-second window for each oscillation event sample.

TABLE I: Distribution of Augmented $PQ|V|f$ Tensors

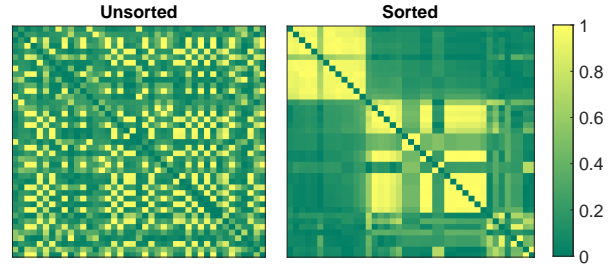| Class	Non-event	Line-event	Generator-event	Oscillation-event
Eastern	720	825	756	708
Western	594	625	666	588

Fig. 4: Weight matrices of the original and the sorted PMU sequence for the 41 PMUs in the Western Interconnection.

TABLE II: F1 Scores for Different Event Classes

| Class	Non-event	Line-event	Generator-event	Oscillation-event
No Sorting	0.858	0.904	0.903	0.941
With Sorting	0.890	0.928	0.932	0.949

be observed that GSP-based PMU sorting arranged highly correlated PMUs into clusters. This approach enables the classifier to better capture the spatial-temporal correlations in the data, thereby yielding more accurate event classification results. To quantify the benefits of the GSP-based PMU sorting, we train the neural network ten times with different random initial parameter and evaluate the performance of each trained neural network on the testing dataset. Table II shows the average F1 score for each event class at the end of training session. Clearly, the proposed GSP-based PMU sorting technique improves the classification performance in terms of F1 scores for all event classes.

C. Benefits of Transfer Learning

In this subsection, we try to illustrate the benefits of transfer learning. First, we train ten neural networks with different random initial parameters using the PMU data from the Eastern Interconnection. We then train ten neural networks on the PMU data from the Western Interconnection with the encoder's initial parameters transferred from that of the ten encoders trained based on the Eastern Interconnection data. Serving as the baseline, another ten neural networks with random initial parameters are also trained using the PMU data from the Western Interconnection. We evaluate the performance of these neural networks on the testing dataset during their training sessions. The average accuracy as well as the 95% confidence intervals are depicted by the red dotted lines and shaded areas in Fig. 5. Clearly, the transfer learning significantly reduces the required training time to reach a desirable event classification accuracy level. In addition, the transfer learning helps shrink
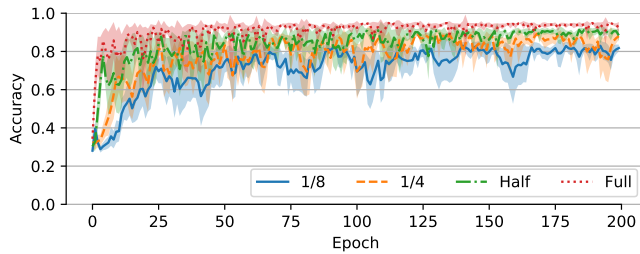
(a) Without transfer learning.
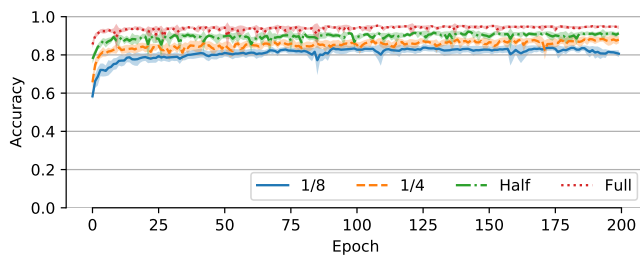
(b) With transfer learning.

Fig. 5: Power system event classification accuracy comparison on the testing dataset of the Western Interconnection given different amount of training data.

the uncertainty bounds considerably, especially during the early stage of training.

A single system operator usually has limited labeled event training data, especially in the initial stage of PMU deployment and operation, which can lead to unstable training behavior for the neural classifier. We investigate how transfer learning could improve the event classification performance with a small amount of training data. To this end, we carry out three sets of experiments and report the testing accuracy with half, $1/4$, and $1/8$ of training data with and without transfer learning. The results are shown in Fig. 5. Without transfer learning, the training sessions become increasingly unstable and slower to converge as the amount of training data decreases. Even with limited training dataset, the transfer learning scheme leads to more stable training sessions with faster convergence rates.

IV. CONCLUSION

This paper proposes a transfer learning framework to improve the applicability of deep neural network-based power system event identification algorithms. This technique is particularly useful for power systems with limited event labels or stringent neural network training time requirement. The core idea is to set the initial values of the target neural network parameters the same as that of a network previously trained on a different power transmission network. The testing results on real-world synchrophasor data gathered from hundreds of PMUs across the Eastern and the Western Interconnection in the U.S. show that the proposed transfer learning approach significantly increases the stability of the neural network training sessions and decreases the required training time.

REFERENCES

[1] D.-I. Kim, T. Y. Chun, S.-H. Yoon, G. Lee, and Y.-J. Shin, "Wavelet-based event detection method using PMU data," *IEEE Transactions on Smart Grid*, vol. 8, no. 3, pp. 1154–1162, Oct. 2015.

[2] S. S. Negi, N. Kishor, K. Uhlen, and R. Negi, "Event detection and its signal characterization in PMU data stream," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3108–3118, Jul. 2017.

[3] S.-W. Sohn, A. J. Allen, S. Kulkarni, W. M. Grady, and S. Santoso, "Event detection method for the PMUs synchrophasor data," in *IEEE Power Electronics and Machines in Wind Applications*, Jul. 2012.

[4] J. Shi, B. Foggo, X. Kong, Y. Cheng, N. Yu, and K. Yamashita, "Online event detection in synchrophasor data with graph signal processing," in *IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Nov. 2020.

[5] Y. Zhou, R. Arghandeh, H. Zou, and C. J. Spanos, "Nonparametric event detection in multiple time series for power distribution networks," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 2, pp. 1619–1628, Jun. 2018.

[6] L. Xie, Y. Chen, and P. Kumar, "Dimensionality reduction of synchrophasor data for early event detection: Linearized analysis," *IEEE Transactions on Power Systems*, vol. 29, no. 6, Nov. 2014.

[7] J. Wu, J. Xiong, P. Shil, and Y. Shi, "Real time anomaly detection in wide area monitoring of smart grids," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov. 2014, pp. 197–204.

[8] Z. Ling, R. C. Qiu, X. He, and L. Chu, "A new approach of exploiting self-adjoint matrix polynomials of large random matrices for anomaly detection and fault location," *IEEE Transactions on Big Data*, May 2019.

[9] X. Shi and R. Qiu, "Dimensionality increment of PMU data for anomaly detection in low observability power systems," *arXiv preprint arXiv:1910.08696*, 2019.

[10] P. Gao, M. Wang, S. G. Ghiocel, J. H. Chow, B. Fardanesh, and G. Stefopoulos, "Missing data recovery by exploiting low-dimensionality in power system synchrophasor measurements," *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1006–1013, Apr. 2015.

[11] Y. Hao, M. Wang, J. H. Chow, E. Farantatos, and M. Patel, "Modelless data quality improvement of streaming synchrophasor measurements by exploiting the low-rank Hankel structure," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6966–6977, Jun. 2018.

[12] J. Shi, N. Yu, E. Keogh, H. K. Chen, and K. Yamashita, "Discovering and labeling power system events in synchrophasor data with matrix profile," in *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. IEEE, Nov. 2019, pp. 1827–1832.

[13] W. Li, M. Wang, and J. H. Chow, "Real-time event identification through low-dimensional subspace characterization of high-dimensional synchrophasor data," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 4937–4947, Jan. 2018.

[14] D. Nguyen, R. Barella, S. A. Wallace, X. Zhao, and X. Liang, "Smart grid line event classification using supervised learning over PMU data streams," in *International Green and Sustainable Computing Conference (IGSC)*, Dec. 2015, pp. 1–8.

[15] O. P. Dahal and S. M. Brahma, "Preliminary work to classify the disturbance events recorded by phasor measurement units," in *IEEE Power and Energy Society General Meeting*, Jul. 2012, pp. 1–8.

[16] W. Wang, H. Yin, C. Chen, A. Till, W. Yao, X. Deng, and Y. Liu, "Frequency disturbance event detection based on synchrophasors and deep learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3593–3605, Feb. 2020.

[17] J. Shi, B. Foggo, and N. Yu, "Power system event identification based on deep neural network with information loading," *IEEE Transactions on Power Systems*, May 2021.

[18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[19] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, Dec. 2014.

[20] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct. 2009.