





Siamese Neural Network Based Few-Shot Learning for Anomaly Detection in Industrial Cyber-Physical Systems

Xiaokang Zhou , Member, IEEE, Wei Liang , Member, IEEE, Shohei Shimizu , Jianhua Ma, Member, IEEE, and Qun Jin , Senior Member, IEEE

I. INTRODUCTION

Abstract—With the increasing population of Industry 4.0, both AI and smart techniques have been applied and become hotly discussed topics in industrial cyber-physical systems (CPS). Intelligent anomaly detection for identifying cyber-physical attacks to guarantee the work efficiency and safety is still a challenging issue, especially when dealing with few labeled data for cyber-physical security protection. In this article, we propose a few-shot learning model with Siamese convolutional neural network (FSL-SCNN), to alleviate the over-fitting issue and enhance the accuracy for intelligent anomaly detection in industrial CPS. A Siamese CNN encoding network is constructed to measure distances of input samples based on their optimized feature representations. A robust cost function design including three specific losses is then proposed to enhance the efficiency of training process. An intelligent anomaly detection algorithm is developed finally. Experiment results based on a fully labeled public dataset and a few labeled dataset demonstrate that our proposed FSL-SCNN can significantly improve false alarm rate (FAR) and F1 scores when detecting intrusion signals for industrial CPS security protection.

Index Terms—Anomaly detection, convolutional neural network (CNN), few-shot learning, industrial cyber-physical systems (CPS), Siamese network.

CYBER-PHYSICAL system (CPS), which can usually be divided into three layers including the physical layer, transmission layer, and application layer, is a multidimensional complex system integrating computation, physical processing, and networking. With the rapid development of Industry 4.0, signals and messages exchanging through networks based on industrial Internet of Things (IIoT) empower the functionality and efficiency of CPS in industrial environments [1], including real-time perception, dynamic control, and information service of large-scale engineering systems. However, the diversity of CPS applications deploying across networks in IIoT makes it vulnerable to both cyber and physical attacks among different levels of systems, especially for message transmissions in smart manufacturing processes.

Currently, due to the new characteristics of different attacks in industrial CPS, it becomes necessary to involve and develop advanced intelligent computing, communication and control technologies to deal with the cyber-physical security issues [2]. Typically, the possibility of industrial CPS compromised by various attacks becomes higher along with the increase of the number of physical sensors and I/O interfaces. For example, in 2015, the Ukrainian State Electric Power Department suffered a malicious code attack and resulted in a power outage, which has been viewed as a typical case of cyber security shortcoming [3]. The openness of modern information and communication technology makes cyberphysical security a significant issue in developing industrial CPS. In particular, intelligent anomaly detection becomes a significant way to identify both cyber and physical attacks among the whole networks for security protection.

Modern AI technologies, including intelligent sensing, smart control, etc., are widely used for behavior monitoring in smart manufacturing. However, there are still several challenges when detecting abnormal signals in industrial CPS. First, the hybrid cyber-physical environment constructed with a cloud infrastructure is a large and complicated distributed system, thus a large volume of industrial data stream (e.g., instruction, accelerometer, video, image, etc.) is generated via a variety of physical systems and sensors. To alleviate the damage caused by malicious attacks in industrial CPS, it requires real-time anomaly detections with high accuracy and timeliness, to facilitate the

Manuscript received June 26, 2020; revised September 16, 2020 and November 26, 2020; accepted December 14, 2020. Date of publication December 31, 2020; date of current version May 3, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2017YFE0117500, Grant 2019YFE0190500, and Grant 2019GK1010, in part by the National Natural Science Foundation of China under Grant 62072171, and in part by the Natural Science Foundation of Hunan Province of China under Grant 2019JJ40150 and Grant 2018JJ2198. Paper no. TII-20-3098. (Corresponding author: Wei Liang.)

Xiaokang Zhou and Shohei Shimizu are with the Faculty of Data Science, Shiga University, Hikone 5228522, Japan, and also with the RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan (e-mail: zhou@biwako.shiga-u.ac.jp; shohei-shimizu@biwako.shiga-u.ac.jp).

Wei Liang is with the Key Laboratory of Hunan Province for New Retail Virtual Reality Technology, Hunan University of Technology and Business, Changsha 410205, China (e-mail: weiliang@csu.edu.cn).

Jianhua Ma is with the Faculty of Computer and Information Sciences, Hosei University, Chiyoda-ku 1028160, Japan (e-mail: jianhua@hosei.ac.jp).

Qun Jin is with the Faculty of Human Sciences, Waseda University, Tokorozawa 3591192, Japan (e-mail: jin@waseda.jp).

Color versions of one or more of the figures in this article are available at <https://doi.org/10.1109/TII.2020.3047675>.

Digital Object Identifier 10.1109/TII.2020.3047675

surveillance of overall performance based on the data stream obtained and transferred from different levels of distributed nodes across the system. Another critical issue in industrial CPS is that such kinds of abnormal events occur rarely in the real world. The low occurrence probability of these anomaly activities results in the lack of well labeled data for model training. In addition, missing of surveillance data, which may be caused by different factors, such as sensor failure, data transferring error, etc., is a common problem in most of industrial systems, but will bring a more difficult situation for automatic data collection and model training toward intelligent anomaly detection. Since conventional learning techniques mainly depend on a large labeled training database, it becomes more challenging when facing the above problems in the real-time surveillance and anomaly detection tasks [4]. Therefore, intelligent strategies need to be designed and developed to deal with the time-consuming issues especially when multiple sensors are used to extract samples based on different frequencies during a more complex data fusion process in industrial CPS [5].

Few-shot learning is an emerging learning paradigm aiming at tackling issues on the lack of training data, which enables models to identify novel categories with only a few sample data provided for them. The key point is that the training sample needs to be carefully selected in order to perfectly match the inference during the testing phase. Each step is designed to simulate a small-sample learning task by subsampling classes and data points (e.g., sampling five classes at one time, each of which is with five labeled samples). To complete the few-shot learning task, a well-trained feature extractor should be devised, and an effective classifier is essential to mine rich information from a small number of labeled samples.

In this article, we propose a Siamese neural network based few-shot learning model to deal with the cyber-physical security protection issue with few labeled data. In particular, a Siamese convolutional neural network (CNN) is constructed to improve the high-dimensional feature learning, and further facilitate the identification of novel classes, based on an optimized relative-feature representation. A robust cost function with three specific losses is designed to enhance the training efficiency, and an improved algorithm is developed for intelligent anomaly detections in industrial CPS. The main contribution of our article can be concluded as follows.

- 1) A few-shot learning model based on Siamese CNN is designed with a relative-feature representation scheme, which can alleviate the over-fitting issue especially when coping with few labeled data in industrial CPS. This model is later referred to as few-shot learning model with Siamese convolutional neural network (FSL-SCNN).
- 2) A robust cost function, considering a combination of the transforming loss in relative-feature representation, the encoding loss during CNN encoding process, and the prediction loss based on the distances between the *anchor* sample and the *positive* and *negative* samples, is introduced, which may significantly enhance the efficiency of training process.
- 3) An intelligent detection algorithm is developed based on a transformed lower dimensional feature representation,

which can be applied for anomaly detections from large amount of industrial CPS data with few labeled samples.

The rest of this article is organized as follows. Section II presents an overview of related works on few-shot learning techniques for malicious attack detections. The proposed FSL-SCNN and anomaly detection implementation are discussed in Section III. Section IV demonstrates the results on performance evaluation for the proposed method. Section V conclude this article.

II. RELATED WORK

In this section, several issues relating to this article in intelligent industrial systems, including analytics on anomaly detection techniques for CPS, and models on few-shot learning in industrial applications, are discussed respectively.

A. Anomaly Detection Techniques for CPS

In current years, researchers have paid great efforts to tackle the vulnerability and security issues for CPS, which were implemented in a variety of applications ranging from data acquisition, surveillance, and industrial control systems. There are various kinds of cyber and physical attacks which may affect the reliability and security of CPS. For example, to explore vulnerabilities in industrial CPS, Alan *et al.* [6] considered a covert attack for service degradation, and introduced a backtracking search optimization algorithm to deal with the system identification attack in cyberphysical control systems. Beg *et al.* [7] focused on the false-data injection attack, and designed a detection framework, to identify changes based on a set of candidate invariants inferred from Simulink/Stateflow diagrams in cyber-physical dc microgrids. To cope with a typical type of denial-of-service (DoS) attacks in CPS, Sun *et al.* [8] proposed a resilient control strategy with a dual-mode algorithm, which could be used for the optimization problem in model predictive control without the consideration of model uncertainties and measurement noises.

Particularly, anomaly detection has been analyzed extensively for purposes of cyber security and system reliability. It becomes crucial to develop appropriate security protection frameworks or control systems to tackle vulnerabilities in industrial CPS under different attack scenarios. Several AI-based approaches have been investigated for cyber-physical security protection including attack identification, fault detection, and tolerant control [9]. Kim *et al.* [10] analyzed the cyber-physical vulnerability, and presented a software-defined networking-based architecture for man-in-the-middle attack. They applied it in a specific communication-based train control system, to improve the resiliency for attack detections. Li *et al.* [11] built a dual deep learning (DL) model with an energy auditing mechanism, to monitor and identify cyber and physical attacks in IoT environments. They designed a disaggregation-aggregation structure to learn the system behaviors for the attack detection. The disaggregation part was used to analyze the energy consumption for cyber-attack identification, and the aggregation part was used to measure the power consumption for physical attack identification. Pearce *et al.* [12] introduced a framework to prevent different kinds

of cyber-physical attacks based on the runtime enforcement, in which the bidirectional timed policies were specified in an industrial CPS application.

Obviously, previous researches have shown the success in applying DL techniques to identify a variety of cyber-physical attacks. However, conventional supervised learning models heavily rely on the prior knowledge and well labeled training samples, which may be difficult in handling the over-fitting issue, and even result in poor performance when detecting new categories with a few samples for anomaly detection in intelligent industrial environments.

B. Few-Shot Learning in Industrial Applications

Few-shot learning is an emerging type of transfer learning technique. By reusing the transferrable knowledge of existing models, a classifier can be built to identify the novel category using only a few labeled training samples [13]. Along with the popularity of DL, few-shot learning model is increasingly drawn attention in modern industrial applications. Gu *et al.* [14] built a recognition network to deal with the few-shot density problem for industrial safety and environmental protection. They employed the model-agnostic meta-learning algorithm to optimize the initial parameters, which could achieve better classification results with only a small number of gradient steps in flare soot applications. Sun *et al.* [15] constructed a feature fusion model based on the so-called focus-area location and high-order integration for few-shot tasks. The few-shot learning was utilized to identify similar regions and extract more discriminative features. Perez-Cabo *et al.* [16] proposed a deep metric learning method for the generalized presentation attack detection problem, in which a triplet focal loss was defined to regularize a new “metric-softmax” loss. They used the few-shot learning to improve the feature representation and distinguish attacks only using the image data. Huang *et al.* [17] developed a few-shot learning model for imbalanced data problems. They designed a gated network structure to analyze the known types and unknown types in anomaly detection, and tested their method in identifying new anomaly types for few-shot learning tasks. Chowdhury *et al.* [18] introduced a DL approach to few-shot intrusion detection. The CNN model, linear support vector machine, and one-nearest neighbor classifier were integrated together in a training model for new feature representations. They argued that their method could be used to identify some minority attack types. Shen *et al.* [19] presented a machine learning-based framework for resource management in wireless communications. The idea of few-shot learning was used in a self-imitation mechanism, which could optimize a new task with a few unlabeled samples based on a pre-trained learning model. Lu *et al.* [20] defined two types of outliers: representation outlier and label outlier, and constructed an attentive profile network model for outlier suppression based on few-shot learning using user-provided data.

Comparing with previous researches, in this article, we construct a few-shot learning model to overcome the overfitting issue, in which a Siamese CNN structure is designed and constructed to alleviate the loss of key features. The proposed model can be applied to enhance the anomaly detection performance for security protection in industrial CPS.

III. FEW-SHOT LEARNING MODEL WITH SIAMESE CNN IN CPS

In this section, we first introduce the system architecture for cyber-physical security protection in industrial CPS. A few-shot learning framework is constructed and presented with a CNN-based Siamese network. An intelligent anomaly detection algorithm is then developed based on a relative-feature representation scheme and a robust cost function design.

A. Problem Definition

Fig. 1 illustrates a typical architecture for security protection in AI-enhanced industrial CPS. Usually, attackers may hack into the CANbus network and send the malicious code to compromise systems. The supervisory control and data acquisition (SCADA) system is involved to monitor and collect signals (e.g., vibration, temperature, and TX&RX packet data) generated across the cyber network, in which the DL-based anomaly detection module is deployed to identify anomalies. Since it is a costly task to collect anomalies with large enough set of samples for traditional model training, a Siamese CNN encoding network model is designed to facilitate the real-time analysis based on few-shot learning, which can improve the intelligent anomaly detection with higher efficiency and accuracy in industrial CPS.

Given an anomaly detection problem in the industrial CPS, two general datasets, D_{nor} and D_{ano} , are taken into considered to indicate the normal and anomaly samples, respectively. $D_{\text{nor}} = \{(x_{\text{nor}_i}, y_{\text{nor}_i}) \mid i = 1, \dots, N_{\text{nor}}\}$, contains N_{nor} labeled normal samples, in which x_{nor_i} is the data sample and y_{nor_i} is the corresponding class label. Likewise, $D_{\text{ano}} = \{(x_{\text{ano}_i}, y_{\text{ano}_i}) \mid i = 1, \dots, N_{\text{ano}}\}$, contains N_{ano} labeled anomaly samples, in which x_{ano_i} is the data sample and y_{ano_i} is the corresponding class label. We assume $N_{\text{nor}} \gg N_{\text{ano}}$ to describe the few-shot learning scenario. Thus a set of samples from D_{ano} is selected to form the support set in each training episode, and the corresponding query set Q , which is used to indicate the unobserved samples of novel classes between different episodes, can be described as $= \{(x_{\text{ano}_j}, y_{\text{ano}_j}) \mid j = 1, \dots, N_q\}$. Summarily, in each episode, we randomly select K malicious attack classes, each of which includes C labeled samples, to form the K -way C -shot learning problem, aiming to enhance the generalization of detection capability of our model especially for novel attack identification.

B. Few-Shot Learning Framework for Anomaly Detection

The proposed FSL-SCNN is designed to tackle the issue on lacking adequate labeled anomaly samples in our detection tasks. Differing from conventional classification models, our FSL-SCNN do not predict the class for an input sample data directly, but calculate the distance between the input samples in terms of their optimized feature representations. In particular, a CNN-based Siamese network is constructed to cope with the few-shot learning problem, thus the novel classes can be identified even with only a few sample data supported. The framework of FSL-SCNN for anomaly detection in CPS is illustrated in Fig. 2.

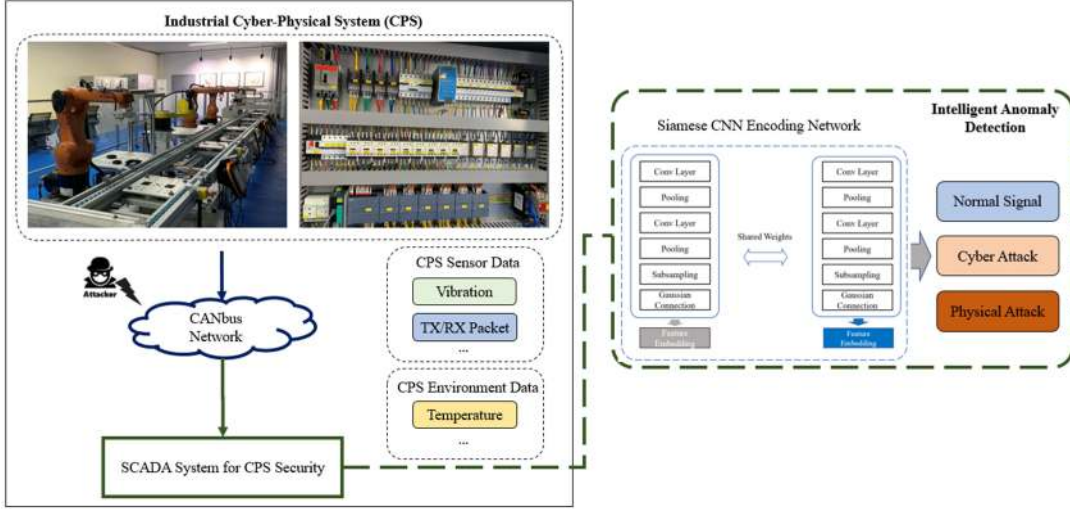


Fig. 1. Typical architecture of security protection for industrial CPS.

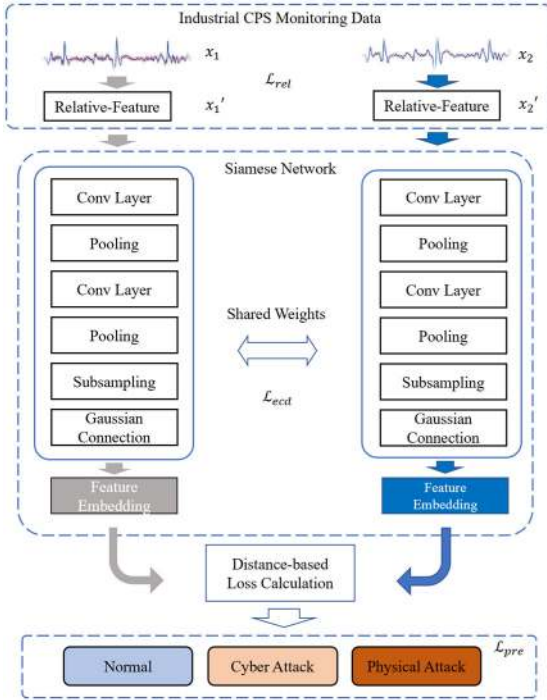


Fig. 2. CNN-based Siamese network for few-shot learning in anomaly detections.

As shown in Fig. 2, to train this DL model, two input sample data (i.e., one from support set and one from query set) for each class will be sent into two identical CNN simultaneously. A relative-feature representation scheme is applied to transform their original features into a lower dimensional representation, which can help the neural network alleviate the overfitting issue, and consequently enhance the detection performance. In the Siamese network, two combinations of convolution layer and pooling layer are introduced to extract feature embeddings. During the testing process, the distance between these two feature

embeddings will be calculated to identify whether these two input samples belong to the same class.

Given x_i as one input sample sent to the FSL-SCNN, the feature embedding $f(x_i)$ extracted by the Siamese CNN can be represented as follows:

$$f(x_i) = \text{CNN}_{ecd}(x_i, \theta_{encoding}) \quad (1)$$

where $\theta_{encoding}$ is the encoding parameter of CNN.

The distance between two feature embeddings from two input samples x_i and x_j is defined and calculated based on the pairwise Euclidean distance, which can be described as follows:

$$D(f(x_i), f(x_j)) = \|f(x_i) - f(x_j)\|^2. \quad (2)$$

Finally, the output of the FSL-SCNN is generated based on the fully connected layer and SoftMax layer, which can be expressed as follows:

$$P(x_i, x_j) = \text{SoftMax}(FC(D(f(x_i), f(x_j)))) \quad (3)$$

where $\text{SoftMax}(\ast)$ indicates the function of SoftMax and $FC(\ast)$ indicates the function of fully connected layer. $P(x_i, x_j)$ represents the probability whether x_i and x_j belong to the same class.

C. Robust Cost Function Design

To ensure the prediction accuracy and false alarm rate (FAR) for anomaly detections from large volume of industrial CPS data with few labeled samples, three losses are considered in our cost function design. As shown in Fig. 2, the transforming loss \mathcal{L}_{rel} is issued in the relative-feature representation. The encoding loss \mathcal{L}_{eed} is generated during the CNN encoding process, which is designed to measure the variance between the transformed relative-features and extracted feature embeddings. The prediction loss \mathcal{L}_{pre} is a triplet loss based on the distances between the anchor sample, and the positive and negative samples.

Considering we only have a few samples in the support set, the dimensionality of original features for input samples

becomes relatively large compared to the total number of support samples, which thus usually leads to the overfitting and poor generalization performance for the model. Motivated by [21], a relative-feature representation for input samples is applied to reduce the dimensionality of original features. Specifically, the transformed features with relatively low dimensionality of an input sample is calculated based on the distance between itself and all the other samples using (2). Given n samples as the input for the model, $\frac{n(n-1)}{2}$ sets of distances need to be calculated. The detailed relative-feature can be calculated and expressed as follows:

$$x_i' = [D(x_i, x_1)^2, D(x_i, x_2)^2, \dots, D(x_i, x_n)^2]. \quad (4)$$

For example, given four samples, x_1, x_2, x_3, x_4 , with their corresponding pairwise distance: $D(x_1, x_2) = 1$, $D(x_1, x_3) = 2$, $D(x_1, x_4) = 3$, $D(x_2, x_3) = 2$, $D(x_2, x_4) = 4$, $D(x_3, x_4) = 1$, the transformed relative-features can be generated in a four-dimensional feature vector as: $x_1' = [0, 1, 4, 9]$, $x_2' = [1, 0, 4, 16]$, $x_3' = [4, 4, 0, 1]$, and $x_4' = [9, 16, 1, 0]$. Obviously, the dimensionality of the input sample can be effectively reduced no matter how large the original sample is.

Therefore, for each input (x_i, y_i) , the loss in relative-feature representation can be defined and calculated as follows:

$$\mathcal{L}_{rel} = \frac{1}{N_q} \sum_{(x_i, y_i)} -\log \left(\frac{\exp(-d(f(x_i), p_m))}{\sum_{m'=1}^K \exp(-d(f(x_i), p_{m'}))} \right) \quad (5)$$

where $d(*, *)$ is the Euclidean distance. p_m is calculated by averaging the samples of class m for relative-feature representations, while $p_{m'}$ is calculated for the corresponding representations in each training episode.

The loss for CNN encoding is employed to measure if there is any loss of key information within the Siamese network. Following the encoding process described in (1), The decoding function based on the Siamese CNN is defined as follows:

$$x_i'' = \text{CNN}_{dcd}(f(x_i), \theta_{decoding}). \quad (6)$$

Since it is difficult to observe the information loss directly during the encoding process, motivated by [22], the relative entropy theory can be used to measure the loss of information based on the real distribution and theoretical distribution. Thus, given a probability distribution of an input sample x_i , the encoding loss is designed to minimize the number of feature embeddings, while retaining the key information of features in the original data. The detailed calculation based on the Kullback–Leibler (KL) divergence can be described as follows:

$$\mathcal{L}_{ecd} = E \left[\sum_{x_i} p(x_i|f(x_i)) \log \left(\frac{p(x_i|f(x_i))}{q(x_i|f(x_i))} \right) \right] \quad (7)$$

where $p(x_i|f(x_i))$ indicates the real distribution of the sample data. $q(x_i|f(x_i))$ is the calculated distribution and can be treated as an approximate to $p(x_i|f(x_i))$.

Furthermore, distances between the *anchor* sample and the *positive* and *negative* samples are considered to measure the prediction loss based on the Siamese network. The detailed

calculation can be formulated as follows:

$$\mathcal{L}_{pre} = \max \left(\left(\frac{D(f(x_i^a), f(x_i^p)) - D(f(x_i^a), f(x_i^n))}{D(f(x_i^a), f(x_i^n)) + \alpha} \right), 0 \right) \quad (8)$$

where x_i^a , x_i^p and x_i^n indicate the *anchor*, *positive*, and *negative* samples respectively. $\alpha \in (0, 1)$ is a coefficient to adjust the FAR in anomaly detection. Usually, it is empirically set as $\alpha > 0.5$ during the training process. The maximum function is used to ensure a minus loss for \mathcal{L}_{pre} , thus the *anchor* sample can be more similar to the *positive* sample than the *negative* one, based on this adversarial design.

D. Intelligent Anomaly Detection for Cyber-Physical Security Protection

To pursue an efficient training performance, the cost function $\mathcal{L}_{FSL-SCNN}$ in the FSL-SCNN is composed based on a combination of the three losses discussed above, which can be defined and expressed as follows:

$$\mathcal{L}_{FSL-SCNN} = \mathcal{L}_{rel} + \tau \cdot \mathcal{L}_{ecd} + \mathcal{L}_{pre} \quad (9)$$

where τ is a balance coefficient to control the encoding loss \mathcal{L}_{ecd} during the training process.

Specifically, $\mathcal{L}_{FSL-SCNN}$ is designed to tackle the following challenges during the few-shot learning process: retaining the critical information when transforming original high-dimensional features into a relatively low dimensionality during the relative-feature representation; and enabling the learning model to present reasonable feature embeddings in the Siamese network, thereby alleviate the overfitting problem when the training data is insufficient. The concrete anomaly detection algorithm is shown in Algorithm 1.

The training process via the proposed FSL-SCNN is divided into three steps: feature transformation, feature encoding and distance comparison. In each training episode, the raw data x is transformed to x_i' based on the relative-feature representation scheme first. x_i' is then formalized into the structured feature embedding $f(x_i')$ through the CNN encoder. According to a selected *Anchor* sample x_i^a , a *positive* sample x_i^p and a *negative* sample x_i^n , the corresponding classes y_i^a , y_i^p , and y_i^n are predicted via the constructed Siamese network respectively. The losses generated during relative-feature representation, CNN encoding, and prediction process, are calculated using the designed cost function as addressed by (5), (7), and (8). Consequently, the model M will be finalized by minimizing the total loss $\mathcal{L}_{FSL-SCNN}$.

IV. EXPERIMENT AND ANALYSIS

In this section, evaluations are conducted to demonstrate the performance of our proposed method for anomaly detection, comparing with other similar mechanisms based on two different datasets.

A. Dataset and Experiment Design

To investigate the effectiveness of the proposed FSL-SCNN, both a fully labeled public dataset and a few labeled dataset are considered in our experiment evaluation. The fully labeled

Algorithm 1: FSL-SCNN Based Anomaly Detection.

Input: A set of anomalies samples $D_{\text{ano}} = \{(x_{\text{ano}_i}, y_{\text{ano}_i}) \mid i = 1, 2, \dots, N_{\text{ano}}\}$
A set of normal signal samples $D_{\text{nor}} = \{(x_{\text{nor}_i}, y_{\text{nor}_i}) \mid i = 1, 2, \dots, N_{\text{nor}}\}$
A set of query samples $Q = \{(x_{\text{ano}_j}, y_{\text{ano}_j}) \mid j = 1, 2, \dots, N_q\}$

Output: A trained anomaly detection model M

- 1: Initialize hyper parameter α , τ , and loss threshold ε
- 2: **while** $\mathcal{L}_{\text{FSL-SCNN}} > \varepsilon$ **do**
- 3: **for** each episode **do**
- 4: Choose k class with c samples from D_{nor} and D_{ano} to build support set
- 5: Choose k class from Q to build query set
- 6: **for** x_i in support set **do**
- 7: Transform x_i into relative representation x_i'
- 8: Calculate transforming loss by Eq. (5)
- 9: Transform x_i' into feature embedding $f(x_i')$ via the CNN Encoder by Eq. (1)
- 10: Calculate encoding loss by Eq. (6)
- 11: Select *anchor* sample x_i^a and predict y_i^a based on $f(x_i^a)$ by Eq. (3)
- 12: Select another *positive* sample x_i^p and *negative* sample x_i^n , predict y_i^p and y_i^n by Eq. (3)
- 13: Calculate prediction loss by Eq. (8)
- 14: Update network to minimize $\mathcal{L}_{\text{FSL-SCNN}}$ by Eq. (9)
- 15: **end for**
- 16: **end for**
- 17: **end while**
- 18: **return** M

public dataset UNSW-NB15, generated by the Australian security laboratory for CPS [23], is applied to evaluate the general prediction performance of the proposed method. This dataset is composed of network traffic packets created using IXIA PerfectStorm tool, including realistic modern normal activity and synthetic contemporary attack behavior packets. It contains nine categories of cyber-physical attacks including: analysis; fuzzers; DOS; generic; backdoor; exploit; reconnaissance; worm; and shellcode. The few labeled dataset used in the experiment is generated in an intelligent CPS for smart manufacturing as illustrated in Fig. 1, in which the network transmission packet is collected via the SCADA system, and contains a small number of randomly generated abnormally high or low transmission rate signals. The average packet amount per second is fluctuated with a normal state of 0.05 KB/s. Specifically, the former dataset is used to evaluate the training efficiency and anomaly detection performance of the proposed method, while the latter one is used to investigate the effectiveness of our method in a cyber-attack scenario.

We selected several classical and widely used machine learning methods, and a Siamese model for anomaly detection in CPS as the baseline methods. Specifically, the time series analysis (TSA) which is introduced as a non-machine learning technique,

classical machine learning methods including Naïve Bayes (NB), random forest (RF), and one-shot support vector machine (OS-SVM), are compared in this article. It is noted that OS-SVM is a kernel-based variation of SVM method with only one-shot data sample for each class, thus is selected to compare with the proposed FSL-SCNN. In addition, a Siamese convolutional autoencoder (SCAE) model [24], comprising twin convolutional autoencoders, is involved for comparison evaluations as well.

Four widely used metrics, precision, recall, F1, and FAR, are applied and calculated according to whether normal/anomaly signals have been identified correctly or not, in order to demonstrate the performances of these mentioned methods based on the fully labeled public dataset. In particular, FAR is an important metric to evaluate the anomaly detection performance in CPS especially in unbalanced dataset. The lower the FAR, the better performance is achieved by the model in practical scenarios.

B. Anomaly Detection Performance Evaluation

We chose stochastic gradient descent (SGD) as the optimizer to train the model. The learning rate was set to 0.1 and we iterated 800 times to investigate the training process in the experiment. The transforming loss, the encoding loss, and the prediction loss obtained in each iteration using UNSW-NB15 are shown in Fig. 3 respectively.

As shown in Fig. 3, the overall performances of the three losses decline fast and become relatively stable. Relatively, the error rates of transforming loss and prediction loss fluctuate greatly during the learning process according to Fig. 3(a) and (c), while the error rate of encoding loss drops sharply and trends to stable after 200 iterations according to Fig. 3(b). This training result indicates the applicability and suitability of our model in few-shot learning.

Furthermore, to evaluate the feature embedding effect based on the relative-feature representation and CNN encoding in the Siamese network, we investigate all the six methods based on the principal components analysis (PCA) result. The visualization comparisons based on UNSW-NB15 are shown in Fig. 4.

The distinct difference in terms of data distributions shown in Fig. 4 demonstrates the imbalance in the dataset, as well as the corresponding features. In other words, the number of normal samples is much more than the number of the attack samples. It can be observed that feature embeddings based on our proposed FSL-SCNN result in a better clustering performance. The better the clustering performance, the better effect of the feature extraction will be. Moreover, comparing with other five methods, the method generates an obvious clustering result with few overlaps among features in two distinguished classes. This result indicates the effectiveness of the combination of relative-feature representation and CNN encoding in reducing dimensionality and retaining the key feature information during the learning process within our Siamese network.

We go further to evaluate the overall performance of anomaly detection, based on precision, recall, F1, and FAR in an imbalanced dataset. Especially, the FAR is a significant indicator to

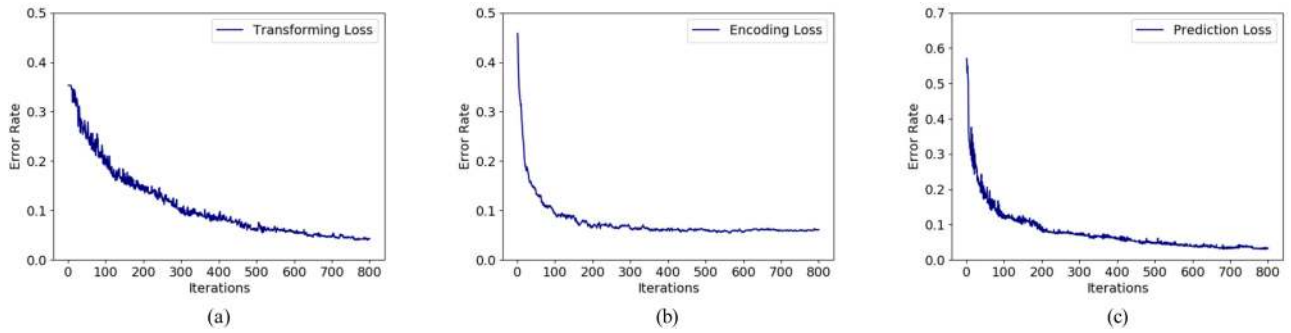


Fig. 3. Evaluation on training efficiency. (a) Transforming loss curve. (b) Encoding loss curve. (c) Prediction loss curve.

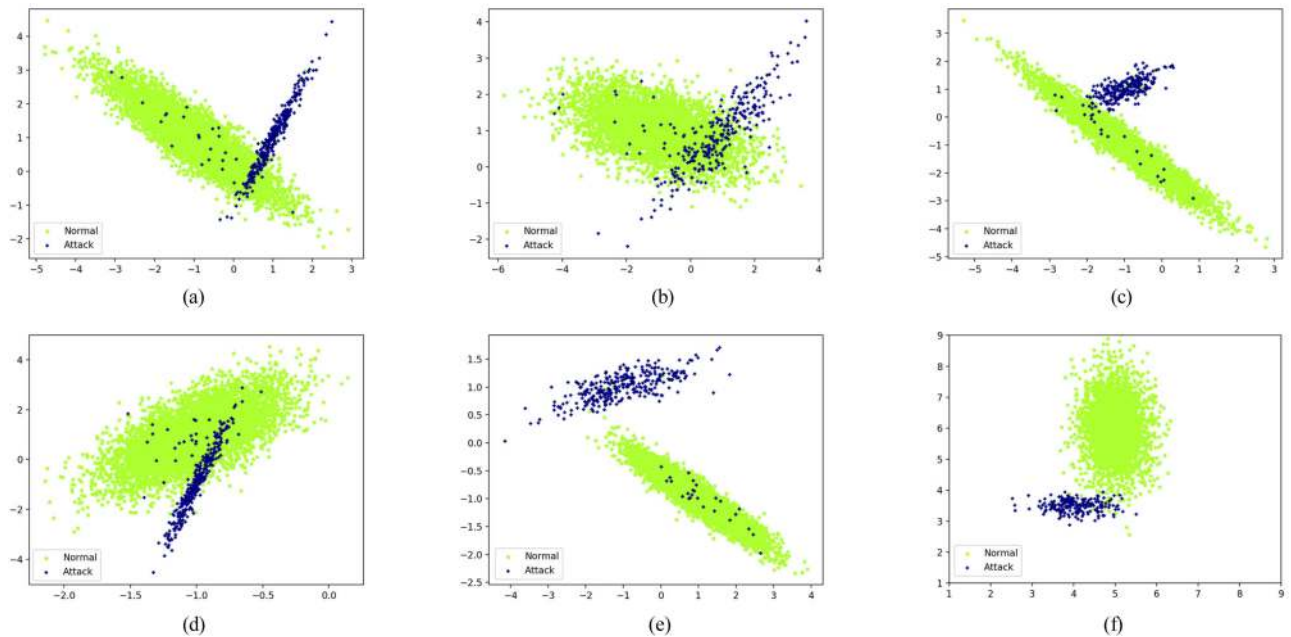


Fig. 4. Feature embedding evaluation based on PCA. (a) TSA. (b) NB. (c) RF. (d) OS-SVM. (e) SCAE. (f) FSL-SCNN.

TABLE I
ANOMALY DETECTION PERFORMANCE COMPARISONS

Methods	Precision	Recall	F1 score	FAR
TSA	0.871	0.870	0.870	0.157
OS-SVM	0.895	0.911	0.903	0.089
RF	0.889	0.821	0.854	0.179
SCAE	0.899	0.910	0.904	0.090
NB	0.746	0.852	0.795	0.168
FSL-SCNN	0.906	0.968	0.936	0.047

demonstrate the performance of anomaly detection in the real world. The results are compared and given in Table I.

According to Table I, we observe that the proposed FSL-SCNN has achieved the best results in F1 score and FAR at 0.936 and 0.047 respectively. Since FAR is an important indicator to evaluate the performance of anomaly detection in CPS as we discussed earlier, this result shows that because of the relative-feature representation scheme and the robust cost function designed in our model, the FSL-SCNN can not only

distinguish the anomaly signals from the normal ones efficiently, but also reduce the false detection rate in the few-shot learning scenario.

In addition, we investigate the effectiveness of the method in terms of anomaly detection in a real-world cyber-attack scenario. The comparison experiment was conducted based on the few labeled dataset collected in a real CPS as illustrated in Fig. 1. We compared the true attacks and detected anomalies according to the network throughput (bytes per second) captured in the CPS. The evaluation results are illustrated in Fig. 5.

As shown in Fig. 5, we observe the true attacks and detected anomalies respectively, based on the continuous signals generated in the CPS across the timeline from 0 to 1600 s. Anomalies are detected via the proposed CNN-based Siamese network. Obviously, it can be viewed as a few-shot learning problem because there are only a few attacks within the timeline, as depicted in Fig. 5(a). Comparing with the detected anomalies in Fig. 5(b), it is found that most of the cyber-attacks have been effectively identified, which indicates the usefulness of the

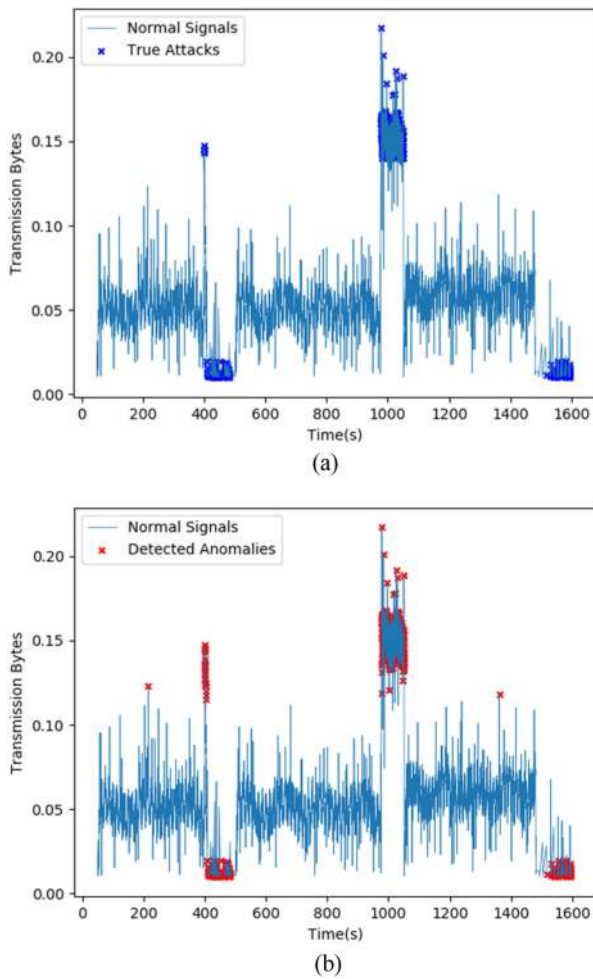


Fig. 5. Cyber-attack analysis based on throughput statistics. (a) True attacks. (b) Detected anomalies.

proposed FSL-SCNN in the real few-shot learning scenario for anomaly detection in industrial CPS.

V. CONCLUSION

In this article, to enhance the cyber-physical security protection in intelligent industrial systems, we proposed the FSL-SCNN to deal with the few labeled and imbalanced dataset generated in industrial CPS for intelligent anomaly detection.

A Siamese CNN encoding network were constructed to measure the distance for input samples based on their optimized feature representations, instead of returning the prediction result directly. The Siamese network structure was capable of identifying novel classes of cyber-physical attacks, even with a few labeled training samples. To alleviate the overfitting issue, a relative-feature representation scheme was utilized to transform original features into a lower dimensional representation. A robust cost function design was introduced, in which three specific losses, including the transforming loss in relative-feature representation, the encoding loss during CNN encoding process, and the prediction loss based on the distances between the

anchor sample, and the positive and negative samples, were seamlessly integrated together to enhance the training efficiency. An intelligent anomaly detection algorithm was then developed to deal with the few labeled data generated in industrial CPS. Experiments and evaluations based on a fully labeled public dataset and a few labeled dataset demonstrated that the method could significantly improve the F1 score and reduce the FAR score comparing with other related methods, which indicated the effectiveness of the proposed model in detecting intrusion signals with few labeled samples in industrial CPS environments.

In future studies, we will go further to conduct more evaluations in different situations to improve the algorithm with better accuracy and efficiency.

REFERENCES

- [1] K. Park, J. Kim, H. Lim, and Y. Eun, "Robust path diversity for network quality of service in cyber-physical systems," *IEEE Trans. Ind. Informat.*, vol. 10, no. 4, pp. 2204–2215, Nov. 2014.
- [2] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-physical systems security—A survey," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 1802–1831, Dec. 2017.
- [3] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security, and privacy issues in Internet-of-Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1250–1258, Oct. 2017.
- [4] T. Wang, J. Xu, W. Zhang, Z. Gu, and H. Zhong, "Self-adaptive cloud monitoring with online anomaly detection," *Future Gener. Comput. Syst.*, vol. 80, pp. 89–101, Mar. 2018.
- [5] R. Sakthivel, S. Santra, and B. Kaviarasan, "Resilient sampled-data control design for singular networked systems with random missing data," *J. Franklin Inst.*, vol. 355, no. 3, pp. 1040–1072, Feb. 2018.
- [6] A. O. de Sá, L. F. R. d. C. Carmo, and R. C. S. Machado, "Covert attacks in cyber-physical control systems," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1641–1651, Aug. 2017.
- [7] O. A. Beg, T. T. Johnson, and A. Davoudi, "Detection of false-data injection attacks in cyber-physical DC microgrids," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2693–2703, Oct. 2017.
- [8] Q. Sun, K. Zhang, and Y. Shi, "Resilient model predictive control of cyber-physical systems under DoS attacks," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4920–4927, Jul. 2020.
- [9] F. Farivar, M. S. Haghighi, A. Jolfaei, and M. Alazab, "Artificial intelligence for detection, estimation, and compensation of malicious attacks in nonlinear cyber-physical systems, and industrial IoT," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2716–2725, Apr. 2020.
- [10] S. Kim, Y. Won, I. Park, Y. Eun, and K. Park, "Cyber-physical vulnerability analysis of communication-based train control," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6353–6362, Aug. 2019.
- [11] F. Li, Y. Shi, A. Shinde, J. Ye, and W. Song, "Enhanced cyber-physical security in Internet of Things through energy auditing," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 5224–5231, Jun. 2019.
- [12] H. Pearce, S. Pinisetty, P. S. Roop, M. M. Y. Kuo, and A. Ukil, "Smart I/O modules for mitigating cyber-physical attacks on industrial control systems," *IEEE Trans. Ind. Informat.*, vol. 16, no. 7, pp. 4659–4669, Jul. 2020.
- [13] M. Uchida, "Human error tolerant anomaly detection based on time-periodic packet sampling," *Knowl. Based Syst.*, vol. 106, pp. 242–250, Aug. 2016.
- [14] K. Gu, Y. Zhang, and J. Qiao, "Ensemble meta learning for few-shot soot density recognition," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, pp. 2261–2270, Mar. 2021.
- [15] X. Sun, H. Xv, J. Dong, H. Zhou, C. Chen, and Q. Li, "Few-shot learning for domain-specific fine-grained image classification," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3588–3598, Apr. 2021.
- [16] D. Pérez-Cabo, D. Jiménez-Cabello, A. Costa-Pazo, and R. J. López-Sastre, "Deep anomaly detection for generalized face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis., Pattern Recognit. Workshops*, 2019, pp. 1591–1600.
- [17] S. Huang *et al.*, "A gated few-shot learning model for anomaly detection," in *Proc. Int. Conf. Inf. Netw.*, 2020, pp. 505–509.

- [18] M. M. U. Chowdhury, F. Hammond, G. Konowicz, C. Xin, H. Wu, and J. Li, "A few-shot deep learning approach for improved intrusion detection," in *Proc. IEEE 8th Annu. Ubiquitous Comput., Electron., Mobile Commun. Conf.*, 2017, pp. 456–462.
- [19] Y. Shen, Y. Shi, J. Zhang, and K. B. Letaief, "LORM: Learning to optimize for resource management in wireless networks with few training samples," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 665–679, Jan. 2020.
- [20] J. Lu, S. Jin, J. Liang, and C. Zhang, "Robust few-shot learning for user-provided data," *IEEE Trans. Neural Netw., Learn. Syst.*, early access, Apr. 2020, doi: [10.1109/TNNLS.2020.2984710](https://doi.org/10.1109/TNNLS.2020.2984710).
- [21] D. Das and C. S. G. Lee, "A two-stage approach to few-shot learning for image recognition," *IEEE Trans. Image Process.*, vol. 29, pp. 3336–3350, 2020.
- [22] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *Statist. Mach. Learn.*, vol. 1, pp. 1–14, 2014.
- [23] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set, and the comparison with the KDD99 data set," in *Proc. Inf. Secur. J., Glob. Perspective*, Jan. 2016, pp. 18–31.
- [24] D. Droghini, S. Squartini, E. Principi, L. Gabrielli, and F. Piazza, "Audio metric learning by using siamese autoencoders for one-shot human fall detection," *IEEE Trans. Emerg. Top. Comput. Intell.*, early access, doi: [10.1109/TETCI.2019.2948151](https://doi.org/10.1109/TETCI.2019.2948151).



Xiaokang Zhou (Member, IEEE) received the Ph.D. degree in human sciences from Waseda University, Tokyo, Japan, in 2014.

He is currently an Associate Professor with the Faculty of Data Science, Shiga University, Hikone, Japan. From 2012 to 2015, he was a Research Associate with the Faculty of Human Sciences, Waseda University. Since 2017, he has been a Visiting Researcher with the RIKEN Center for Advanced Intelligence Project, RIKEN, Japan. He has been engaged

in interdisciplinary research works in the fields of computer science and engineering, information systems, and social and human informatics. His recent research interests include ubiquitous computing, big data, machine learning, behavior and cognitive informatics, cyber-physical-social-system, and cyber intelligence and security.

Dr. Zhou is a Member of the IEEE Computer Society, and Association for Computing Machinery, USA, Information Processing Society of Japan, and Japanese Society for Artificial Intelligence, Japan, and China Computer Federation, China.



Wei Liang (Member, IEEE) received the M.S. and Ph.D. degrees in computer science from Central South University, Changsha, China, in 2005 and 2016, respectively.

From 2014 to 2015, he was a Researcher with the Department of Human Informatics and Cognitive Sciences, Waseda University. Currently, he is with the Key Laboratory of Hunan Province for Mobile Business Intelligence, Hunan University of Technology and Business, Changsha China. He has authored or coauthored more than 20

papers at various conferences and journals. His research interests include information retrieval, data mining, and artificial intelligence.

Dr. Liang is a Member of the IEEE Computer Society, and China Computer Federation, China.



Shohei Shimizu received the Ph.D. degree in statistical science engineering from Osaka University, Osaka, Japan, in 2006.

He is currently a Professor with the Faculty of Data Science, Shiga University, Hikone, Japan and leads the Causal Inference Team, RIKEN Center for Advanced Intelligence Project. His research interests include statistical methodologies for learning data generating processes such as structural equation modeling and independent component analysis and their applica-

tion to causal inference.

Dr. Shimizu was the recipient of Hayashi Chikio Award (Excellence Award) from the Behaviormetric Society in 2016. Since 2016, he has been a Coordinating Editor of Springer Behaviormetrika.



Jianhua Ma (Member, IEEE) received the Ph.D. degree in information engineering from Xidian University, Xi'an, China, in 1990.

He is a Professor with the Department of Digital Media, Faculty of Computer and Information Sciences, Hosei University, Japan. Since 1996, he has been one of pioneers in research on Hyper World and Cyber World). He first proposed Ubiquitous Intelligence towards Smart World, which he envisioned in 2004, and was featured in the European ID People Magazine in

2005. He has conducted several unique CW-related projects including the Cyber Individual (Cyber-I), which was featured by and highlighted on the front page of IEEE Computing Now in 2011. He has founded three IEEE Congresses on "Smart World," "Cybermatics," and "Cyber Science and Technology," respectively, as well as IEEE Conferences on Ubiquitous Intelligence and Computing, Pervasive Intelligence and Computing, Advanced and Trusted Computing, Dependable, Autonomic and Secure Computing, Cyber-Physical and Social Computing, Internet of Things, and Internet of People. He has authored or coauthored more than 300 papers, co-authored five books and edited over 30 journal special issues. His research interests include multimedia, networking, pervasive computing, social computing, wearable technology, IoT, smart things, and cyber intelligence.

Dr. Ma is a Member of the IEEE Computer Society and Association for Computing Machinery, the Chair of IEEE SMC Technical Committee on Cybermatics, the founding chair of IEEE CIS Technical Committee on Smart World, and in the advisory board of IEEE Computer Society Technical Committee on Scalable Computing.



Qun Jin (Senior Member, IEEE) received the Ph.D. degree in electrical engineering and computer science from Nihon University, Tokyo, Japan, in 1992.

He is a Professor with the Networked Information Systems Laboratory, Department of Human Informatics and Cognitive Sciences, Faculty of Human Sciences, Waseda University, Tokyo, Japan. He has been extensively engaged in research works in the fields of computer science, information systems, and social and human informatics.

He seeks to exploit the rich interdependence between theory and practice in his work with interdisciplinary and integrated approaches. His recent research interests include human-centric ubiquitous computing, behavior and cognitive informatics, big data, data quality assurance and sustainable use, personal analytics and individual modeling, intelligence computing, blockchain, cyber security, cyber-enabled applications in healthcare, and computing for well-being.

Dr. Jin is a Senior Member of the Association of Computing Machinery, IEEE Computer Society and Information Processing Society of Japan.