Fast and Accurate Deep Learning Framework for Secure Fault Diagnosis in the Industrial Internet of Things

Youcef Djenouri, Asma Belhadi, Gautam Srivastava, Uttam Ghosh, Pushpita Chatterjee†, and Jerry Chun-Wei Lin*

***Abstract*—This paper introduced a new deep learning framework for fault diagnosis in electrical power systems. The framework integrates the convolution neural network and different regression models to visually identify which faults have occurred in electric power systems. The approach includes three main steps, data preparation, object detection, and hyper-parameter optimization. Inspired by deep learning, evolutionary computation techniques, different strategies have been proposed in each step of the process. In addition, we propose a new hyper-parameters optimization model based on evolutionary computation that can be used to tune parameters of our deep learning framework. In the validation of the framework's usefulness, experimental evaluation is executed using the well known and challenging VOC 2012, the COCO datasets, and the large NESTA 162-bus system. The results are very promising against many current state-of-the-art solutions in terms of runtime and accuracy performances.**

***Index Terms*—Genetic algorithm, Chinese news mining, trading strategy, technical indicators, expected fluctuation analysis.**

I. Introduction

The Industrial Internet of Things (IIoT), as well as Industry 4.0, connect devices to industrial machines, processes, as well as workers using them across a multitude of industrial use cases like manufacturing, logistical supply chain, energy systems, transportation, and healthcare. There has been a recent emergence of promising alternatives using deep learning in processing vast amounts of data (big data), from which knowledge can be extracted. The use of deep learning and AI-based systems in IIoT (Industrial Internet of Things) and/or Industry 4.0, more specifically in electric power systems has been gaining traction in the last ten years [1], [2], in particular, numerous computer vision systems [3], [4] have been implemented in electric power systems environments. Object detection [5]–[7] has come into fruition as a popular research area in IIoT and electrical systems, where it can be noted that the aim is to identify objects from electric images. This line of study will continue in the same direction, and a new intelligent algorithm is proposed to efficiently, and accurately identify fault diagnosis on electric power systems.

Y. Djenouri is with Department of Mathematics and Cybernetics, SINTEF Digital, Oslo, Norway, youcef.djenouri@sintef.no

A. Belhadi is with Department of Technology, Kristiania University College, Oslo, Norway, asma.belhadi@kristiania.no

G. Srivastava is with Department of Mathematics & Computer Science, Brandon University, Canada, and Research Centre for Interneural Computing, China Medical University, Taichung, Taiwan, srivastavag@brandonu.ca

Uttam Ghosh is with Vanderbilt University, 5718 Nashville, Tennessee, United States, 37221, uttam.ghosh@vanderbilt.edu

Pushpita Chatterjee is with the Future Networking Research Group, Ton Duc Thang University, Ho Chi Minh City, Vietnam and the Faculty of Electrical and Electronics Engineering, Ton Duc Thang University, Ho Chi Minh City, Vietnam, Vietnam, pushpita.c@ieee.org (†Co-corresponding author)

J. C. W. Lin is with Department of Computer Science, Electrical Engineering and Mathematical Sciences, Western Norway University of Applied Sciences, Bergen, Norway, jerrylin@ieee.org (*Corresponding author)

Manuscript received April 30, 2020.

A. Motivations

Solutions to AI-based fault diagnosis systems in electrical power systems [8]–[11] are known to be high in time complexity coupled with chronic low accuracy. Also, region-based solutions [12] obtain much better efficiency as compared to single-pass solutions [6]. The processes suffer from further problems with high runtime and less precision. The explanation for this deterioration is that the previous past initiatives all needed to accomplish complex models with a high number of parameters to be checked. One of the key reasons that these solutions have a major challenge is that they have to construct extremely complicated structures with a high number of parameters to be fixed. There have been high success rates of late using data mining, and deep learning [13]–[15]. These methods merge the benefits of data mining methods which allow the discovery of the relevant knowledge from the data, and the benefits of convolution neural networks in learning the machine learning tasks from visual features. Following this trend of success, the research presented here in this work is an end-to-end framework, which explores data mining to extract the most relevant features for fault diagnosis, and a deep learning model to detect different faults in electric power systems. This reduces the time usually incurred by processing the current solution. The main limitation of the end-to-end framework is that a high number of hyper-parameters need to be tuned. Therefore, efficient hyper-parameters optimization techniques should be adopted. Thus, our end-to-end framework benefits by utilizing meta-heuristics in tuning parameters of deep learning models [16]–[18], this research work incorporates both evolutionary computation to tune the parameters of our deep learning model. When direct comparisons are made to previous works, the work before you makes use of several novel innovations that will be shown to improve both the training as well as inference speed showing an increase in detection accuracy.

B. Contributions

In this work, the RCNN+ framework is proposed with an expectation to create a group of efficient learning models for fault diagnosis in electric power systems. It is an end-to-end framework, which incorporates feature selection, deep learning, as well as evolutionary computation approaches. The electric images database is first pre-processed using feature selection, next to deep learning is used to detect the fault diagnosis. Evolutionary computation is integrated into the deep learning model to optimize its hyper-parameters. Holding these facts as notes, our key contributions to this paper can be concluded as follows:

1) We develop a novel feature selection mode that can be used to deduct the unnecessary features from the database of the image, the set of features are extracted using the sift extractor, we then select the most relevant features using the greedy search by optimizing the diversification function among the electric images data.
2) An accurate object detection model is proposed by adopting the Fast-RCNN algorithm [5] on electric images data. We integrate hard negative mining, feature concatenation, and multi-scale training to detect fault diagnosis on electric power systems.
3) We propose a new hyper-parameters optimization model to refine the used parameters of our deep learning algorithm. This model is inspired by the evolutionary computation approaches.
4) We examine RCNN+ by thoroughly evaluating its computational time and accuracy with baseline object detection and fault/error diagnostic approaches. We used different image databases: the challenging VOC 2012, COCO datasets, and the real large NESTA 162-bus system. This evaluation shows that RCNN+ outperforms the baseline algorithms in both runtime and accuracy.

II. Related Work

Solutions to fault diagnosis in electric power systems differ from the nature of the method used in the detection process. Some algorithms use evolutionary and uncertainty computation, some methods use either machine learning and/or deep learning methods while other methods use hybrid models. Saura [9] was the first to make use of the fault diagnosis approach which is based on fault signature which has been observed to be linked to an output voltage of rectifiers. It makes detecting fault diagnosis allowable using a collection of phase-shifting transformer configurations in the most well-known and common fault scenarios. Wang *et al.* [8] proposed a non-dominant sorting in GA which is used to solve the well-known fault diagnosis problem. The fault diagnosis problem has always been considered as a distinct multi-objective optimization problem where we use the Pareto approach to assist in solving the problem. Wang *et al.* [19] investigate uncertainty in fault diagnosis of many power systems. The authors propose a neural system that is defined as interval-valued fuzzy spiking. This system that uses interval-valued fuzzy logic is amalgamated along with spiking neural systems to represent uncertainty within a power system. Similarly, Zhang *et al.* [20] proposed an uncertainty model for sensor faults detection problem by designing a residual generator and analyzing its quantitative influence on sensor faults. Ding [12] introduced a DCN, a short form for a deep convolutional network, where the use of wavelet packet energy images was made for input in the spindle bearing fault diagnosis. To be able to fully find a distinct hierarchical representation, a multi-scale layer is built upon right after the final convolutional layer, which can concatenate outputs of the final convolutional layer with the previous pooling layer. Althobiani [21] introduced a novel scheme for fault diagnosis in reciprocating compressor valves. The use of deep-belief networks with the gaussian visible units is introduced. The scheme makes use of a hierarchical structure with many restricted Boltzmann machines that are stacked as well as a greedy layer-by-layer learning algorithm. Chen *et al.* [22] introduced the use of distributed fault diagnosis for the multi-machine environment based on deterministic learning theory. A learning estimator is first created for each machine in the network to accumulate the local fault. A distributed fault diagnosis model is then trained to monitor the power system of all machines and identify the global fault diagnosis from the local fault of each machine. Zheng *et al.* [23] developed the stochastic hybrid automata approach to identify and detect fault diagnosis. It processes simultaneously many of the continuous variables which include charge and voltage states as well as the discrete dynamics which include faulty as well as normal modes. Wang *et al.* [24] proposed a hybrid auto-encoder deep network with principal component analysis and support vector machine to identify the fault diagnosis in power systems. Thus, instead of using a softmax classifier, a support vector machine classifier with a Gaussian kernel is integrated into the deep model. Principle component analysis is also used to accurately pre-process the data before the classification stage.

It should be evident from this short literature review that solutions to fault diagnosis algorithms for electric power systems suffer from the detection rate due to many reasons, i) Some methods use the whole data features in the detection process, these approaches suffer from the computational time. ii) Some methods use traditional machine learning approaches, which suffer from accuracy. iii) A variety of deep learning models require a lot of tuning parameters, and it is not easy to refine the hyper-parameters of any of these deep learning models. Motivated by the success of feature selection, deep learning, and evolutionary computation in solving complex problems [25]–[29], in the next section, we propose a framework that is defined as end-to-end, which combines feature selection, CNN, and hyper-parameters optimization, to examine the electric images data, and accurately identify the fault diagnosis from electric power systems efficiently.

III. RCNN+ Framework

A. Principle

In this section, the RCNN+ framework is developed to identify faults and anomalies in the industrial internet of things environments. The applicability of the proposed framework on the electric power system is given in the experimentation

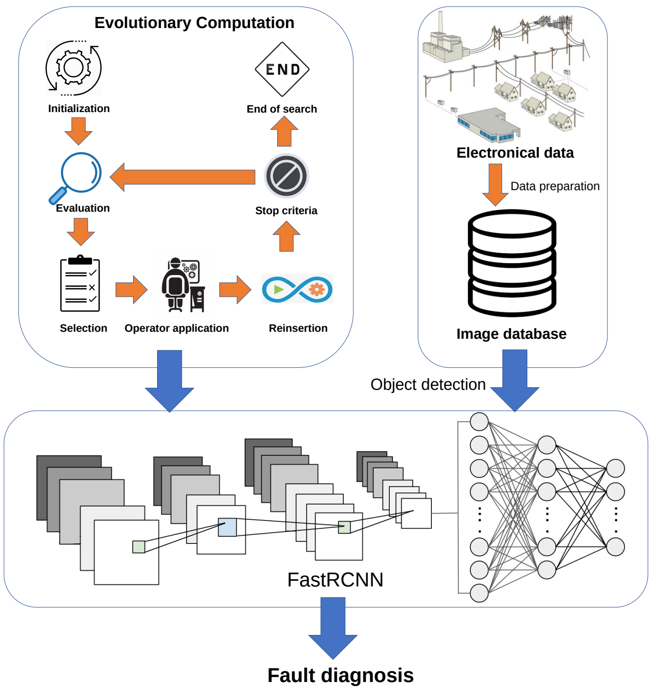

Fig. 1: The developed RCNN+ framework

section. RCNN+ integrates deep learning, data mining, evolutionary computation. From the designed model shown in Fig. 1, RCNN+ consists of three stages: i) Data Preparation: This step integrates methods of image processing, and feature selection for cleaning, and preparing the data to the object detection model. Note that the data is collected using long-range wireless communication technologies such as 4G and 5G. ii) Object Detection: a faster RCNN algorithm is adopted to the electric power system data to identify fault diagnosis, and iii) Hyper-parameters Optimization: Evolutionary computation algorithms are used to learn the different hyper-parameters of the RCNN+ model.

B. Data Preparation

In reality, the images of electric power systems could be as high as $10,000$ or $100,000$ pixels [30]. From the processed results, a huge number of region proposals (e.g., millions or billions) are then produced, which require a very high computational cost and huge memory usage in the entire system. In certain cases, the system will be suddenly blocked after several days and weeks of processing. To tackle this well-known problem and limitation, we develop and implement a pre-processing strategy to prune and filter the number of pixels. We used the image processing step to extract both local, and global features from the set of images, then we integrate the feature selection to figure out the relevant features for the object detection process. It can be summarized as the following two steps:

STEP 1 – SIFT extractor: SIFT extractor [31] is used to derive the set of key points within the scale space of a given image I.

First, The scale-space function, $L(I, \sigma)$, is defined in Equation 1.

$$L(I, \sigma) = G(I, \sigma) * I, \tag{1}$$

where $G(I, \sigma)$ is the Gaussian kernel, and $*$ is the convolution operator.

Second, we determine the position of each candidate keypoint using the interpolation process. It calculates the location that is interpolated right to the extremum. This methodology improves both the stability of the solution as well as matching. The Taylor function $D(I, \sigma)$ is made for interpolation, which is given in Equation 2.

$$D(I, \sigma) = D + \frac{dD^T}{dI} I + 0.5 I^T \frac{d^2 D}{dI^2} I \tag{2}$$

Third, we generate a vector which is called a descriptor for every key point. We create a distinct set of orientation histograms on $4 * 4$ pixel neighborhoods making use of eight bins each. The descriptor vector as defined earlier then becomes every vector of all histograms.

STEP 2 – Feature Selection: The main focus of feature selection is to reduce the number of features by using an optimization function. More formally, given the features of each image I, note $\mathcal{F}_I$, the aim is to select the set of features $\mathcal{F}'_I$ such as $\mathcal{F}'_I \subseteq \mathcal{F}_I$. Our feature selection is based on diversification criteria. We assume that fault diagnosis on electric power systems does not appear often in the same frame, and if they appear, they should close to each other. Consequently, the features of each image should be diversified among the different pixels of such an image. More formally, the optimization function used during the feature selection process is given in Equation 3.

$$\arg\max_{\mathcal{F}'_I} \sum_{i=1}^{|\mathcal{F}'_I|} \sum_{j=1}^{|\mathcal{F}'_I|} Distance(\mathcal{F}'^{(i)}_I, \mathcal{F}'^{(j)}_I) \tag{3}$$

To solve Equation 3, we used the greedy search algorithm. The process starts by generating the initial solution represented by all SIFT features of the image. We delete one feature from the initial solution and evaluate its associated solution using Equation 3. We repeat the process as described until a max number of iterations is achieved, or no improvement is observed. At each step of the algorithm, we only keep the best solution which maximizes the Equation 3.

C. Object Detection

This step aims to identify the fault diagnosis from the input image data. We inspire by the Fast-RCNN principle, which is considered the state-of-the-art object detection solutions [5]. Fast-RCNN can be mainly processed as the following steps:

1) **Region Proposal Determination**: This step aims to compute the regions of interest, the potential regions are represented by bounding boxes, that might be the object allocated. The classical RCNN generates a high number of bounding boxes per image, which yields the overall process high and memory time-consuming. Fast-RCNN

[5] uses a more efficient method by using the convolution neural network to determine the bounding boxes. The neural network is launched to propose bounding boxes by using the ground truth of the training images.

2) **Fast-RCNN**: This step aims to classify regions of images into objects, and refining the boundaries of those regions. Both classification and regression are used.

In this paper, a Fast-RCNN is extended for fault diagnosis on electric power systems. First, the CNN model is trained on the Fast-RCNN making use of the transfer learning process. We train our Fast-RCNN on ImageNet dataset [1] and then use the pre-trained model to the fault diagnosis dataset. In this step, we generate a hard negative, which enriches the network in the training procedure. The combination between feature concatenation and multi-scale training is applied that can be used to speed up the performance of the described trained model. A detailed explanation of our adaptation is given as follows:

1) **Feature Concatenation:** Faster RCNN network performs the regions of interests pooling only on the final feature map layer which assists in the generation of features within the region. This strategy is incomplete and misses some important features, and consequently, the accuracy performance is decreased. To address this issue, different level features are combined with features maps of multiple convolution layers. Multiple feature maps' polling result is concatenated and re-scaled using L2 normalization that assists in the generation of the final pooling features which are used for detection tasks.

2) **Hard Negative Mining**: This strategy aims to identify hard negatives, regions where the network makes an error prediction. Hard negatives are entered into the network using reinforcement learning to enhance the performance of the developed approach. Hard negatives are harvested from the second iteration of our training process, where a region is considered as hard negative if its intersection over union over the ground truth-region is no greater than 40%.

3) **Multi-Scale Training**: The classical Fast-RCNN uses a fixed scale for generating the bounding boxes. In real-world applications such as electric power systems, objects to be detected are multi-scales. Different scales are used to generate the bounding boxes, in this work, we consider five different scales of bounding boxes, (tiny, small, medium, large, and big) to capture objects with different sizes. Thus, five different groups are created, each group consists of bounding boxes of the same size. In this context, the region proposal determination is launched for each group of bounding boxes. At the end of this step, we merge the generated bounding boxes to the convolution neural network for classification and regression steps.

4) **Feature Concatenation**: Faster RCNN network performs the regions of interests pooling only on the final feature map layer which assists in the generation of features within the region. This strategy is incomplete and misses some important features, and consequently, the accuracy performance is decreased. To address this issue, different level features are combined with features maps of multiple convolution layers. Multiple feature maps' polling result is concatenated and re-scaled using L2 normalization that assists in the generation of the final pooling features which are used for detection tasks.

[1] http://www.image-net.org/

D. Hyper-parameters Optimization

The purpose of this part is to determine an optimal set of hyper-parameters of the RCNN+ algorithm. We define the set of hyper-parameters $\mathcal{HP} = \{\mathcal{HP}_1, \mathcal{HP}_2, \dots, \mathcal{HP}_r\}$, where r is the number of hyper-parameters of the RCNN+ algorithm. Each $\mathcal{HP}_i$ is represented by the set of possible values of this hyper-parameter. We define the configuration space $\mathcal{C}$ by the set of all possible configurations, where each configuration represents a vector of possible values of all the hyper-parameters in $\mathcal{HP}$. The hyper-parameters optimization problem aims to derive the optimal configuration which gives the best accuracy in both classifications and regression rates. The size of the configuration space depends on the number of all possible values of the hyper-parameters, and it is determined as given in Equation 4.

$$|\mathcal{C}| = \prod_{i=1}^{r} |\mathcal{HP}_i| \tag{4}$$

The configuration space is considerably huge, for instance, if we only consider $1,000$ possible values for epoch parameter, 100 possible value for error rate and $1,000$ possible values for the number of bounding boxes, the size of the configuration space is 100 million configurations. Therefore, exhaustive search methods are not suitable to solve such a problem. To deal with this issue, evolutionary computation approaches are used. We then present the detailed components of the developed solution as follows.

1) Population Initialization: The initial population of pop_size individuals should be distributed among the configuration space $\mathcal{C}$. This allows exploration of different configurations and covers most regions of the configuration space $\mathcal{C}$. The initial population is generated by respecting diversity, the process starts by generating randomly an individual represented by one configuration in $\mathcal{C}$. From this individual, we generate $pop\_size - 1$ individuals, where each new individual could be dissimilar to the already generated individuals. The dissimilarity between the two configurations is determined using the distance between the configurations of these individuals. The initial population, noted $\mathcal{P}$, should maximize the diversification function described as given in Equation 5.

$$Diversify(\mathcal{P}) = \sum_{i=1}^{pop\_size} \sum_{j=1}^{pop\_size} Distance(\mathcal{C}_i, \mathcal{C}_j), \tag{5}$$

where $Distance(\mathcal{C}_i, \mathcal{C}_j)$ is the distance between the configurations of the i^{th}, and j^{th} individuals, respectively.

2) Crossover: To generate new offspring, the following steps are applied in each of two individuals of the current population:

- We generate a crossover point randomly ranges from 1 to r, which splits each individual into two parts, *left side*, and *right side*.
- The left side of the first individual is transferred to the left side of the first offspring and the right side of the first individual is copied to the right side of the second offspring.
- The left side of the second individual is copied to the left side part of the second offspring and the right side of the second individual is copied to the right side of the first offspring.

3) Mutation: The mutation operation stimulates the diversification search. The technique we use consists of altering the value of one parameter of each existing configuration randomly. The mutation point is generated randomly which is ranges from 1 to r. We iteratively update the value of the mutation point of each configuration in the offspring generated by the crossover operator.

4) Local Search: The local search operator starts from an individual and then recursively moves to its neighbors. The neighborhood is defined by updating the value of one hyper-parameters in the current configuration. This process is repeated for all individuals in the population, and the maximum number of iterations.

5) Fitness Function: As mentioned above, RCNN+ can be used to maximize both the classification and regression ratio. Thus, we use a multi-objective function to evaluate the individuals of the populations as follows:

$$Fitness(\mathcal{C}_i) = \alpha \times CR_{RCNN+}(\mathcal{C}_i) + \beta \times RR_{RCNN+}(\mathcal{C}_i) \quad (6)$$

Note that,

- $\mathcal{C}_i$ is the configuration of the i^th individual in the population.
- $CR_{RCNN+}(\mathcal{C}_i)$: is the classification ratio of the RCNN+ algorithm by applying $\mathcal{C}_i$.
- $RR_{RCNN+}(\mathcal{C}_i)$: is the regression ratio of the RCNN+ algorithm by applying $\mathcal{C}_i$.
- α, and β are two user parameters set between 0.0, and 1.0.

Based on these operations, we proposed two algorithms as follows for hyper-parameters optimization. The first one is based on the mimetic algorithm, and the second one on bees swarm optimization.

6) Mimetic Algorithm: First, the initial population size which is defined as pop_size is randomly generated. Every individual is then built based on the population initialization. Next, local search operators as well as mutation, and crossover are applied that are useful in the generation of configurations from $\mathcal{C}$. To maintain consistent population size, every individual is evaluated making use of the fitness function and focus is placed on keeping the first good-quality pop_size individuals. All others are removed at this stage. This entire process is then repeated in multiple iterations until the max number of iterations is reached.

7) Bees Swarm Optimization Algorithm: Good features are found using an initial bee that settles to find a strong configuration. Through this initial configuration, a distinct set of configurations are determined known as the *SearchArea* of the larger search space. This is done using Equation 5. Every bee considers from the *SearchArea* a configuration as a starting point. Once the local search processing is accomplished, every bee then will communicate what they consider as the best configuration visited all other neighboring bees. This process is completed using a table known as *Dance*. During the next iterations, one configuration from *Dance* will then become the reference configuration. To ensure that know cyclic iterations occur, a taboo list is created of previous reference configurations. Quality criteria are used to choose each reference configuration. That being said, after a set amount of time if the swarm itself as a whole sees that the reference configurations are not improving, a criteria diversification process is utilized to avoid becoming trapped in a local optimum which does not provide any benefit globally. The taboo list is used to create the diversification criteria locating the further past reference configuration from the current one. Once the optimal configuration is located or the maximum iterations (variable) are reached, the algorithm ceases.

IV. Performance Evaluation

The performance evaluation of the proposed framework (RCNN+) is confirmed through experimental evaluation. The VOC 2012 is used, a standard image database, as well as COCO, and a real NESTA162-bus data. The specifics of those databases are given below:

1) **VOC 2012**: The VOC 2012 images database is then used in the experiments for the performance evaluation, which has 17.125 images for varied objects. The images were of very high resolution and were greater than 200x200 pixels per image. In general, there are 20 classes in this database, and each class is considered as a single of the objects, e.g., bird, dining table, person, and among others [32].
2) **COCO** [33]: We used the challenging COCO images database, which contains 83.000 images for varied objects. Each of these images is generated by a large number of pixels, which are used in high-quality image processing (e.g., more than 200x200 pixels of each image). There are about 80 classes in the COCO database.
3) **NESTA162-bus Data** [34]: it is a set of data that contains N-1 possible contingencies which represent all plausible operating points for any given energy demand profile. This set contains over 1 million points. In this dataset topology changes are also included for N-1 investigations.

To examined the observed objects, computing runtime, and accuracy represented by mAP (mean Average Precision) are conducted and verified. mAP is used to test object detection systems, which can be defined and denoted as:

$$mAP = \frac{\sum_{i=0}^{n} AvgP(i)}{n}, \quad (7)$$

where n is considered as the detected objects among all objects, and $AvgP(i)$ is calculated as the precision results at i-rank. For example, the first i-ranked object is then taken into the consideration but ignored others.

The implemented models are performed on a machine fitted with an Intel-Core i7 processor and combined with NVIDIA GeForce GTX 1070 GPU. To simulate IoT environment, the ZigBee[2] system is used in the whole process. To accurately assess the training phase, the parameters are defined by the evolutionary computation (EC) model. The optimized RCNN+ is then evaluated with the up-to-date object detection solutions under a varied number of images in databases, as well as the number of detected images.

A. Parameters Setting

The main purpose of the conducted experiment is to refine the parameters of the RCNN+ model. Fig. 2 presents the convergence of the hyper-parameters optimization algorithms by using a different number of bees for the bees swarm optimization and a different number of generations for the mimetic algorithm. From the results, we can see that the RCNN+ reaches convergence of 85% on VOC 2012, 84% for COCO, and 88% for NESTA162-bus data. As RCNN+ includes different steps, feature selection, objection, and hyper-parameters optimization, in the following, the parameters setting of each step of RCNN+ is explained:

1) **Feature selection**: Greedy algorithm is used in this step, it requires a maximum number of iterations as parameter setting. Therefore, the number of maximum iterations is varied from 1 to 100. The best value of this parameter is used in the next experiment.
2) **Object Detection**: Adopted FastRCNN algorithm is used in this step, which requires a high number of parameters. The parameter setting of this step is automatically performed by the hyper-parameters optimization step. Therefore, the hyper-parameters optimization algorithm is responsible to refine the parameters of FastRCNN including the parameters used in the region determination, the classification, and regression stages. We used selective search for region determination, which requires several bounding boxes per image as a parameter. We varied this parameter from 500 to $2,000$. We used a convolution neural network in the classification stage. We varied the number of epochs from 100 to $1,000$, and the error learning rate from 0.001 to 0.009. We used the support vector machine in the regression stage. We varied the epsilon rate from 0.001 to 0.009. The best values of these parameters are used in the next experiment.
3) **Hyper-parameters Optimization**: This step first needs to select the best method between the evolutionary computation strategy, and then select the best parameters of each method. If the mimetic algorithm is selected, then the population size, the number of generations, the mutation rate, the crossover rate, and the number of neighbours should be well selected. We varied the population size, and the number of generations from 1 to 100, respectively, the mutation and the crossover rates from 0.1 to 0.9, respectively, and the number of neighbours from 1 to 10. If the bees swarm optimization algorithm is considered, then the number of bees, the number of iterations, and the number of neighbours should be well selected. We varied the number of bees, and the number of iterations from 1 to 100, respectively, and the number of neighbours from 1 to 10. The best values of these parameters are used in the next experiment.

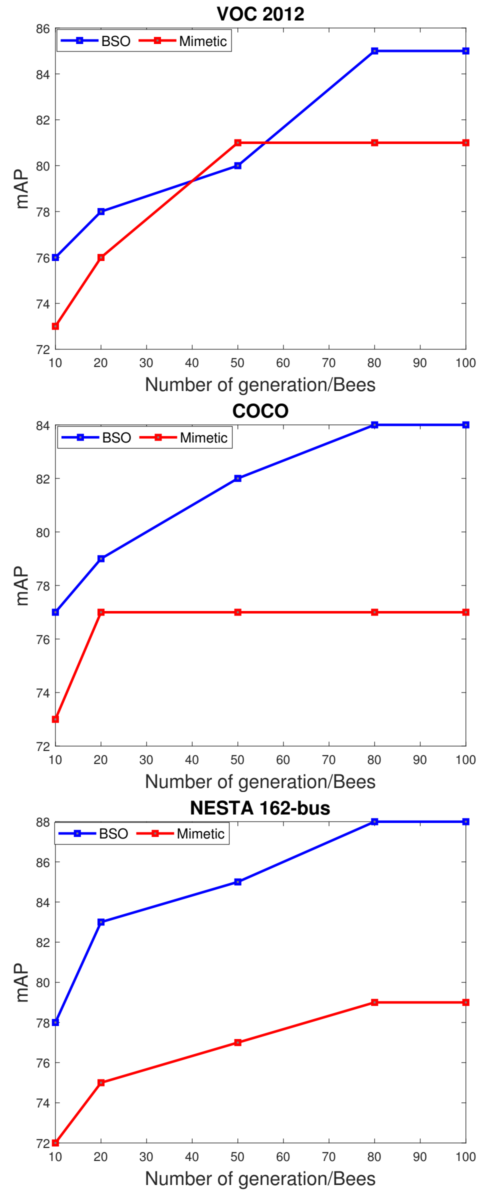

Fig. 2: Convergence of the RCNN+ by varying different hyper-optimization algorithms (BSO: Bees Swarm Optimization), and Mimetic Algorithm, and on different datasets **VOC 2012**, **COCO** and **NESTA162-bus data**.

The best values of the parameters setting step are reported in Table I.

[2]https://zigbeealliance.org/

TABLE I: Best Parameters of the RCNN+.

Steps	Parameters	`VOC 2012`	`COCO`	`Nesta162-bus`
Feature Selection	Number of Iterations of	85	80	73
Object Detection	Number of Bounding Boxes	750	1,100	1,200
	Number of Epochs	250	350	500
	Error Learning Rate	0.002	0.004	0.005
	Epsilon Rate	0.003	0.005	0.004
Hyper-parameters Optimization	Method	Mimetic	Bees Swarm Optimization	Bees Swarm Optimization
	Population Size/Number of Bees	75	50	60
	Number of Generations/ Number of Iterations	55	60	65
	Mutation Rate	0.5	-	-
	Crossover Rate	0.6	-	-
	Number of Neighbors	5	7	9

B. RCNN+ Vs State-of-the-art Object Detection Algorithms

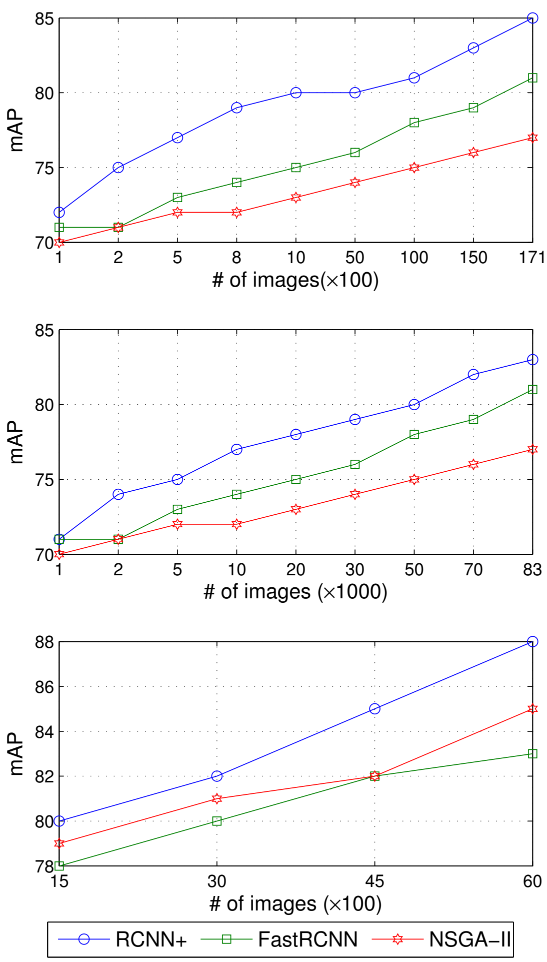

Fig. 3: Accuracy between the RCNN+ v.s. the state-of-the-art fault diagnosis algorithms. The used datasets from the top to down are respectively **VOC 2012**, **COCO** and **NESTA162-bus data**.

The experimental evaluation aims to compare RCNN+ with two baseline algorithms, namely Fast-RCNN [5], and NSGA-II [8] in terms of accuracy and runtime. Fig. 3 present the accuracy of the RCNN+ approach on `VOC 2012`, `COCO`, and `NESTA162-bus databases`, compared with Fast-RCNN [5], and NSGA-II [8]. By experimenting with how many images are used as the input data, RCNN+ achieves the best

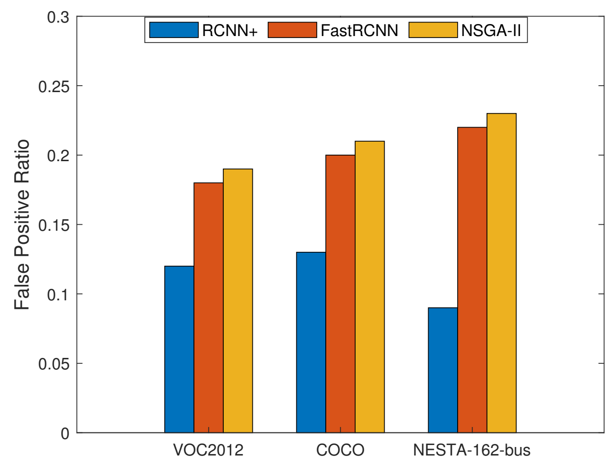

Fig. 4: False positive ratio between the RCNN+ v.s. the state-of-the-art fault diagnosis algorithms. The used datasets from the top to down are respectively **VOC 2012**, **COCO** and **NESTA162-bus data**.

performance compared to the two state-of-the-art algorithms in terms of mAP value as the accuracy result. In addition, Fig. 4 shows the superiority of the RCNN+ in terms of false-positive ratio. This is explained by the fact that the RCNN+ uses efficient strategies to extract the relevant features using the feature selection, the learning is highly optimized using the feature concatenation, the hard negative mining, and multi-scale training. In addition, the parameters are well selected using hyper-parameters optimization. Fig. 5 present the computational time processing between the RCNN+ and the state-of-the-art fault diagnosis algorithms Fast-RCNN [5], and NSGA-II [8] using well-known datasets **VOC 2012**, **COCO** and **NESTA162-bus data**. By varying with the number of image queries from 1 to $10,000$ queries, RCNN+ outperforms the two baseline algorithms in terms of runtime. This is explained by the fact that the RCNN+ explores a new and complete methodology for fault diagnosis problems based on feature selection. This methodology allows to reduce the number of bounding boxes, and therefore, reduce the whole computational process.

V. Discussion and Future Perspectives

In addition to distinguishing objects from the image database, the proposed system examines the numerous corre-

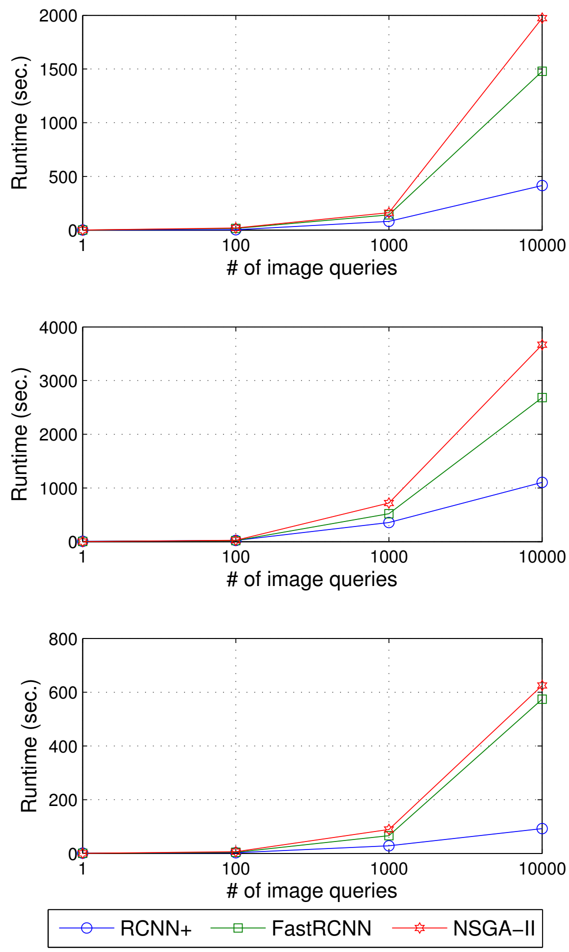

Fig. 5: Runtime between the RCNN+ v.s. the state-of-the-art fault diagnosis algorithms. The used datasets from the top to down are respectively **VOC 2012**, **COCO** and **NESTA162-bus data**.

lations and similarities between images and detects anomalies from electric power data. We claim in respect of object detection that given the object detection and the techniques of hyper-optimization, the fault diagnosis can be detected rapidly and precisely. RCNN+ is an example of how various approaches can be integrated to improve the learning process from an in-depth analysis perspective. In the designed model, we employ feature concatenation, hard negative mining, feature selection, and evolutionary computation to explore electric power systems data. This adaptation takes place at various phases including the elimination of noises, discover the related clusters, learning processes and optimizing the hyper-parameters. The research also finds that the learning process benefits from pre-processing data through the use of feature selection. This helps to accelerate the learning process by generating powerful and coherent models, and each model is learned from clean images. The last point is that, compared to other algorithms, the designed model is more generic and could be extended to more than a certain computer vision problem; the topic of object detection in this paper is an example to show how our framework is to be implemented. The developed system can handle other computer vision problems such as classification, etc.

As far as future work is concerned, the strong results garnered in this paper may lead to different directions investigated later on:

1) **Data Reduction**: In practical scenarios, the number of energy images is too huge, where the fault diagnosis objects to be detected varied in type. Feature selection is one of the powerful data reduction techniques for 2D image analysis [35]. Statistically speaking, it is hard to select the best features of such images since they are computed from the convolution operators. To well reduce the data used in the detection model, a decomposition strategy could be used. The aim is to divide the database of the image into different clusters, and then create different models, one for each cluster of similar images. *k*-means is one of the methods used to group datasets as several homogeneous and similar clusters, and is most widely used in image datasets. Additional techniques can be used to enhance the clustering process and then reduce the number of features shared by the images of different clusters. The incorporation into the RCNN+ system of various clustering strategies such as partitioning, intelligent hierarchical, overlapping, or other research fields, such as entity resolution and/or record linkage, is an important subject for future work. It is also possible to find a suitable method for automatically setting the number of clusters as a constant value. It is not very efficient to use many runs for revealing the best number of clusters. Alternative progress is to build a knowledge-based model having each training image database. After that, the correlation between the meta-features (e.g., number of features, number of images, and luminous pixel values) of the image databases is then examined, as well as the best number of clusters. This helps to estimate the best number of clusters of the database of new images automatically.
2) **Improving the learning step**: We aim to improve the performance of RCNN+ by using high-performance computing resources, such as GPUs, supercomputers and cluster computing for more advanced computer vision applications in the electrical power system environment. This paper aims to build independent work for each image cluster in compliance with high-performance computing issues such as divergence of threads, synchronization, communication, memory management, as well as load balancing. Also, it is necessary to have efficient strategies to process the load balancing issue. One solution for this limitation is to implement the clustering strategies that can figure out the equitable clusters by considering the number of images of each cluster. An alternative way is to design new strategies for repairing clusters and to figure out the clusters under the consideration of a similar number of images. It is also interesting to utilize the RCNN+ on the MapReduce

platform that can be used to improve both the training step, as well as the inference step.

3) **Case studies**: A case study is already shown in this work for an RCNN+ application on electric system data. Gauging the strength of the results from the first case study, RCNN+ can be extended in the future to solve more complex problems from specific domains that are needing learning frameworks for big data. As an example, medical data, as well as intelligent transportation systems, may prove to be perfect applicable scenarios for RCNN+ or any of its future derivatives. Another potential future use of RCNN+ is with sensor-generated data, in real-world, real-time systems like the Internet of Things (IoT) as well as Cyber-Physical Systems (CPS). Some examples include traffic management and other IoT scenarios including smart grids and green technologies. Here, the process for learning needs to be done within short periods, often in real-time with limited latency. Another important case study is to deal directly with 3D images, in this context, we need to adopt the graph convolution neural network [36] to learn the 3D images, for instances the 3D point clouds data.

VI. Conclusion

This work introduces a new deep learning framework for fault diagnosis in electric power systems. It combines the convolution neural network and different regression models to visually identify which faults occurred in the electric power systems. The approach includes three main steps, data preparation, object detection, and hyper-parameters optimization. The feature selection is used to pre-process the image database and extract the most relevant features of each image. An extended version of the faster RCNN algorithm is used to detect fault diagnosis, integrating transfer learning, feature concatenation, hard negative mining, and multi-scale training. The overall process is optimized by the evolutionary computation algorithms, in which the best hyper-parameters of the trained model are retrieved. The experimental results of the designed model are very promising against NSGA-II, and FastRCNN in terms of computational time, and the mean average precision of the detected objects in VOC 2012, COCO, and the NESTA 162-bus datasets. As a perspective, we plan to extend the proposed framework by targeting data reduction, other deep learning models, more promising case studies on industrial informatics.

References

[1] L. Yin, Q. Gao, L. Zhao, B. Zhang, T. Wang, S. Li, and H. Liu, “A review of machine learning for new generation smart dispatch in power systems,” *Engineering Applications of Artificial Intelligence*, vol. 88, p. 103372, 2020.

[2] G. Sideratos, A. Ikonomopoulos, and N. D. Hatziargyriou, “A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks,” *Electric Power Systems Research*, vol. 178, p. 106025, 2020.

[3] J. Wu, J. Liu, J. Ma, K. Chen, and C. Xu, “Classification of power loads based on an improved denoising deconvolutional auto-encoder,” *Applied Soft Computing*, vol. 87, p. 105959, 2020.

[4] J. Wang, L. P. Tchapmi, A. P. Ravikumar, M. McGuire, C. S. Bell, D. Zimmerle, S. Savarese, and A. R. Brandt, “Machine vision for natural gas methane emissions detection using an infrared camera,” *Applied Energy*, vol. 257, p. 113998, 2020.

[5] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.

[6] H. Chen, Z. He, B. Shi, and T. Zhong, “Research on recognition method of electrical components based on yolo v3,” *IEEE Access*, vol. 7, pp. 157 818–157 829, 2019.

[7] H. Yang, L. Chen, M. Chen, Z. Ma, F. Deng, M. Li, and X. Li, “Tender tea shoots recognition and positioning for picking robot using improved yolo-v3 model,” *IEEE Access*, vol. 7, pp. 180 998–181 011, 2019.

[8] S. Wang, D. Zhao, J. Yuan, H. Li, and Y. Gao, “Application of nsga-ii algorithm for fault diagnosis in power system,” *Electric Power Systems Research*, vol. 175, p. 105893, 2019.

[9] J. Saura, M. Bakkar, and S. Bogarra, “Open-circuit fault diagnosis and maintenance in multi-pulse parallel and series tru topologies,” *IEEE Transactions on Power Electronics*, 2020.

[10] T. Wang, X. Wei, J. Wang, T. Huang, H. Peng, X. Song, L. V. Cabrera, and M. J. Pérez-Jiménez, “A weighted corrective fuzzy reasoning spiking neural p system for fault diagnosis in power systems with variable topologies,” *Engineering Applications of Artificial Intelligence*, vol. 92, p. 103680, 2020.

[11] V. H. Ferreira, R. Zanghi, M. Z. Fortes, S. Gomes Jr, and A. P. A. da Silva, “Probabilistic transmission line fault diagnosis using autonomous neural models,” *Electric Power Systems Research*, vol. 185, p. 106360, 2020.

[12] X. Ding and Q. He, “Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 8, pp. 1926–1935, 2017.

[13] Y. Djenouri, G. Srivastava, and J. C. W. Lin, “Fast and accurate convolution neural network for detecting manufacturing data,” *IEEE Transactions on Industrial Informatics*, 2020.

[14] Y. Djenouri, A. Belhadi, J. C.-W. Lin, D. Djenouri, and A. Cano, “A survey on urban traffic anomalies detection algorithms,” *IEEE Access*, 2019.

[15] D. Djenouri, R. Laidi, Y. Djenouri, and I. Balasingham, “Machine learning for smart building applications: Review and taxonomy,” *ACM Computing Surveys (CSUR)*, vol. 52, no. 2, pp. 1–36, 2019.

[16] R. Vinayakumar, M. Alazab, S. Srinivasan, Q. V. Pham, S. K. Padannayil, and K. Simran, “A visualized botnet detection system based deep learning for the internet of things networks of smart cities,” *IEEE Transactions on Industry Applications*, 2020.

[17] T. S. Rao, S. T. Ram, and J. Subrahmanyam, “Neural network with adaptive evolutionary learning and cascaded support vector machine for fault localization and diagnosis in power distribution system,” *Evolutionary Intelligence*, pp. 1–12, 2020.

[18] L. Du, Z. Cui, L. Wang, and J. Ma, “Structure tuning method on deep convolutional generative adversarial network with nondominated sorting genetic algorithm ii,” *Concurrency and Computation: Practice and Experience*, p. e5688, 2020.

[19] J. Wang, H. Peng, W. Yu, J. Ming, M. J. Pérez-Jiménez, C. Tao, and X. Huang, “Interval-valued fuzzy spiking neural p systems for fault diagnosis of power transmission networks,” *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 102–109, 2019.

[20] J. Zhang, P. D. Christofides, X. He, Z. Wu, Z. Yinghong, and D. Zhou, “Robust detection of intermittent sensor faults in stochastic ltv systems,” *Neurocomputing*, 2020.

[21] V. T. Tran, F. AlThobiani, and A. Ball, “An approach to fault diagnosis of reciprocating compressor valves using teager–kaiser energy operator and deep belief networks,” *Expert Systems with Applications*, vol. 41, no. 9, pp. 4113–4122, 2014.

[22] T. Chen, D. J. Hill, and C. Wang, “Distributed fast fault diagnosis for multimachine power systems via deterministic learning,” *IEEE Transactions on Industrial Electronics*, 2019.

[23] C. Zheng, Z. Chen, and D. Huang, “Fault diagnosis of voltage sensor and current sensor for lithium-ion battery pack using hybrid system modeling and unscented particle filter,” *Energy*, vol. 191, p. 116504, 2020.

[24] Y. Wang, M. Liu, Z. Bao, and S. Zhang, “Stacked sparse autoencoder with pca and svm for data-based line trip fault diagnosis in power systems,” *Neural Computing and Applications*, vol. 31, no. 10, pp. 6719–6731, 2019.

[25] Z. Liu, M. Zhang, F. Liu, and B. Zhang, “Fault diagnosis on the braking system of heavy-haul train based on multi-dimensional feature fusion

and gbdt enhanced classification," *IEEE Transactions on Industrial Informatics*, 2020.
[26] Z. Zhen, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, X. Lu, T. Wang, and Y. Lin, "Deep learning based surface irradiance mapping model for solar pv power forecasting using sky image," *IEEE Transactions on Industry Applications*, 2020.
[27] M. E. Khoda, T. Imam, J. Kamruzzaman, I. Gondal, and A. Rahman, "Robust malware defense in industrial iot applications using machine learning with selective adversarial samples," *IEEE Transactions on Industry Applications*, 2019.
[28] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 3057–3065, 2017.
[29] M. Baza, N. Lasla, M. Mahmoud, G. Srivastava, and M. Abdallah, "B-ride: Ride sharing with privacy-preservation, trust and fair payment atop public blockchain," *IEEE Transactions on Network Science and Engineering*, 2019.
[30] A. Ryu, H. Ishii, and Y. Hayashi, "Battery smoothing control for photovoltaic system using short-term forecast with total sky images," *Electric Power Systems Research*, vol. 190, p. 106645, 2021.
[31] J. Markel, "The sift algorithm for fundamental frequency estimation," *IEEE Transactions on Audio and Electroacoustics*, vol. 20, no. 5, pp. 367–377, 1972.
[32] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, no. 2, pp. 303–338, 2010.
[33] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *European conference on computer vision*. Springer, 2014, pp. 740–755.
[34] F. Thams, A. Venzke, R. Eriksson, and S. Chatzivasileiadis, "Efficient database generation for data-driven security assessment of power systems," *IEEE Transactions on Power Systems*, 2019.
[35] S.-H. Wang, D. R. Nayak, D. S. Guttery, X. Zhang, and Y.-D. Zhang, "Covid-19 classification by ccshnet with deep fusion using transfer learning and discriminant correlation analysis," *Information Fusion*, vol. 68, pp. 131–148, 2021.
[36] S. H. Wang, V. V. Govindaraj, J. M. Górriz, X. Zhang, and Y. D. Zhang, "Covid-19 classification by fgcnet with deep feature fusion from graph convolutional network and convolutional neural network," *Information Fusion*, vol. 67, pp. 208–229, 2021.