Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 25, Issues 2, pp. 840–847, DOI 10.12694/scpe.v25i2.2486

OPTIMIZATION OF NONLINEAR CONVOLUTIONAL NEURAL NETWORKS BASED ON IMPROVED CHAMELEON GROUP ALGORITHM

QINGTAO ZHANG*

Abstract. In order to solve the most difficult problem of the architectural model established by CNN in solving specific problems, which results in parameter overflow and inefficient training, an optimization algorithm for nonlinear convolutional neural networks based on improved chameleon swarm algorithm is proposed. This article mainly introduces the use of Chameleon Swarm Optimization (PSO) algorithm to research the parameters of CNN architecture, solve them, and achieve the optimization of the optimization model.Although the number of parameters that need to be set up in CNN is very large, this method can find better testing space for Alexnet samples with 5 different images. In order to improve the performance of the improved pruning algorithms, two candidate pruning algorithms are also proposed. The experimental results show that compared with the traditional Alexnet model, the improved pruning method improves the image recognition ability of the Caffe primary parameter set from 1.3% to 5.7%. Conclusion: This method has wide applicability and can be applied to most neural networks which do not require any special functional modules of the Alexnet network model.

Key words: Deep learning, Convolutional neural network, Chameleon group optimization algorithm, Image recognition

1. Introduction. The optimization problem, which dates back to the ancient extremum problem, is a branch of computational science and is now a widely studied topic. Optimization problem requires that the maximum or minimum value of the objective function can be obtained by reasonable search method under certain constraints. Since the target space of the optimization problem is generally huge, it is impossible to use the exhaustive method to solve it. It is necessary to design a suitable optimization method to solve it.

Traditional optimization methods include simple form method, common gradient method, Newton method, etc. Because these optimization methods generally require the objective function to be differentiable, and need to search the search space on a large scale, the algorithm is feasible in theory but not in practice. However, the actual optimization problems are often complex, with the characteristics of non-differentiable, nonlinear and multi-extremum. It is obvious that the traditional optimization method can not meet the requirements of calculation accuracy and convergence speed. Therefore, designing efficient algorithms to solve complex optimization problems has always been a research hotspot in computational science.Optimization problem has always been an important problem to be solved in the field of scientific research and engineering. It has played an important role in the development of the history of science and the progress of human civilization, which makes the study of optimization theory become a very active field. With the deepening of human's understanding and research of the natural world, the scale of the problems involved is also growing, and the large-scale optimization problem becomes an urgent problem to be solved effectively, which raises the requirement for optimization theory. Some optimization theories and algorithms have been further developed due to the increasing performance requirements and the continuous improvement of computing performance. At the same time, these optimization technologies have been successfully applied to a number of practical engineering fields, and achieved certain development, such as industrial production control task scheduling intelligent system [13, 5].

Convolutional neural network is a deep pre crushing neural network, whose basic characteristics are obtained by continuous clustering of extraction methods. It is composed of an access layer, a hidden layer, an entire link layer, and an output layer, which are connected by a solution layer and a sub-layer.

The convolution layer deals with the resolution function of the object image, and extracts the feature of the image. The weight is the same as the window, first in the horizontal translation, then the bottom, so the

*Department of Computer and Information Engineering, Hebei Petroleum University of Technology, 067000, China (QingtaoZhang7@163.com)

operation of the picture is the operation resolution.

In recent years, deep neural networks have become a phenomenon research hotspot due to their remarkable performance advantages, and have been applied to various fields and made remarkable achievements. Deep neural network has greatly exceeded the traditional algorithm architecture in terms of performance, namely, the way of manual features and classifiers, and has been favored by experts and scholars in various fields. The development of neural network began in the late 19th century, and it has been more than a century. From the original MP model to the perceptron model; However, in the trough period, because the feasibility of using multi-layer structure to extend the perceptron model cannot be verified theoretically, the neural network shows certain limitations when dealing with nonlinear problems.

In a sense, the training process of neural network is also the process of dealing with large-scale optimization problems, that is, looking for network parameters that make the model adapt to the data. The mainstream approach to deal with optimization problems on neural network is to use error back propagation (Error Back Propagation, BP). In addition, parameters are updated in the form of error gradient descent. This traditional optimization method needs to calculate the gradient of each parameter in the face of the huge number of parameters in the deep network, which increases the difficulty of solving and requires very high computing power, and the current equipment cannot meet its fast solution.

2. Literature Review. Convolutional neural network (CNNs), as one of the most important depth models, has good feature extraction and generalization ability. It has achieved great success in image processing, target tracking and detection, natural language processing, scene classification, face recognition, audio retrieval, medical diagnosis and many other fields. On the one hand, the rapid development of high-resolution neural network is due to the significant improvement in computer performance, which makes the construction and training of large-scale networks not restricted by the hardware level. On the other hand, due to the development of large-scale data processing, the general scalability of the network is enhanced.

Convolutional neural networks (CNN) is an important method in image recognition, including image resolution technique, clustering technique, and composite layers. Pop, C. B et al. puts forward a model of AlexNet, which is the first time that CNN is superior to the traditional mathematics model. Based on the LeNet-5 model, it is proposed to extend and deepen the network, and to improve the recognition capability of the model. These ideas have received the approval of the scientists and the CNNS with complex, multi - and multi - constraints [14].

Appropriately increasing the scale of network model and training data is helpful to improve the final recognition effect of neural network, but it is bound to be accompanied by a huge amount of computation and long training time. Therefore, the acceleration of convolutional neural networks is now the focus of research. The main acceleration modes focus on the adjustment training algorithm and parallel acceleration. Parallel acceleration mainly uses hardware environment and parallel computing framework to accelerate hardware. FPGA implementation of convolutional neural network has appeared as early as the mid-1990s, which uses arithmetic methods with low accuracy to replace all multipliers. In recent years, more and more studies have been conducted on using FPGA to accelerate convolutional neural network. In addition, Bell LABS implemented ANNA's chip in the early 1990s, which was also the first time to accelerate convolutional neural networks through hardware. In recent years, the Institute of Computing Science of the Chinese Academy of Sciences has proposed deep learning processors DianNao Computer and DaDianNao large computer. Aiming at the underlying hardware of the computer, the operations of each layer of deep learning are integrated into a hardware unit, which can well improve the efficiency of convolutional neural networks [9].

The CNN model proposed in this paper through parameter optimization of the improved chameleon group algorithm has better recognition accuracy than the standard CNN model and can be verified. The method proposed in this paper is suitable for most existing CNN architectures [12].

3. Research Method.

3.1. Chameleon algorithm.

3.1.1. The main idea of the algorithm Chameleon. Chameleon is a hierarchical clustering algorithm that uses qualitative models. In the Chameleon clustering method, two clusters are merged if the intersections and computations between them correspond to the intersections and computations of items in the cluster. The way it works is to put the data items into many small sub-clusters from a shared image, and then combine the sub-classes with a hierarchical clustering algorithm to see the actual results. A unified process model helps detect natural or homogeneous groups and can be applied to any type of data as long as the features are similar. The Chameleon algorithm takes into account cluster connectivity and computation, especially the intrinsic properties of clusters, to identify similar subclusters [10, 18].

3.1.2. Chameleon algorithm. The Chameleon algorithm defines its properties as a k-nearest neighbor graph. Each K-point in the nearest neighbor graph represents a data object, and if data A is one of the k-closest objects of data B, then objects A and B are edges. The nearest image K-concept is dynamically obtained. Community: The electrical community of an object is determined by the density of its siblings. The idea of K-community is expressed dynamically: the local electricity of an object is determined by the density of the place where the object is. In densely populated areas, the definition of community is narrow. In the distribution of objects section, the definition of groups is broader and the area density is denoted as edge weight. Therefore, the edges of a dense object have more weight than the edges of a diffuse object.

3.1.3. The determination of similarity between clusters in the Chameleon algorithm. The Chameleon determines the similarity between clusters by the relative interconnection RI $C_i, C_j)$ and the relative approximation RC $C_i, C_j)$ of two clusters.Chameleon.

(1) Relative interconnection RI $C_i, C_j)$ is defined as the absolute interconnection between C_i and C_j and the normalization of the internal interconnection of two clusters, i.e., the following formula (3.1):

$$\mathrm{RI}\left(C_i, C_j\right)=\frac{\left|E_{C_i, C_i}\right|}{\frac{1}{2}\left|E_{C_i}\right|+\left|E_{C_j}\right|)} \tag{3.1}$$

EC_{C_i,C_i} is the truncated edge of the cluster containing C_i and C_j classified into C_i and C_j; EC_{C_i} (or EC_{C_j}) is the size of the minimum truncated bisector (that is, the weighted sum of the edges that need to be cut off to divide the graph into two roughly equal parts)

(2) Relative approximation RC $C_i, C_j)$ is defined as the normalization of the absolute approximation between C_i and C_j about the internal approximation of the two clusters, namely, the following equation (3.2):

$$\mathrm{RC}\left(C_i, C_j\right)=\frac{S_{\mathrm{EC}\left(C_1, C_j\right)}}{\frac{\left|C_i\right|}{\left|C_i\right|+\left|C_j\right|} S_{\mathrm{EC}}+\frac{\left|C_j\right|}{\left|C_i\right|+\left|C_j\right|} S_{\mathrm{EC}_{C_j}}} \tag{3.2}$$

$S_{\mathrm{EC}(C_1,C_j)}$ is the average weight of the edges connecting vertices C_i and C_j, S_{EC}, $S_{\mathrm{EC}_{C_j}}$ are the average weights of the edges of the minimum truncated bisector of C_i and C_j, respectively.

3.2. Parameter to be optimized. Related issues that need to be optimized when solving neural network problems are the size of the convolutional kernel and the size and type of weighting layers for each feature parameter. convolutional windows network. Parameters are looked up using high-valued floating-point values, then balanced, keeping the desired number of objects and taking into account the parameters to fall back to the range if they are lost outside of the dynamic configuration. This is because if the step size is optimized by the improved chameleon swarm algorithm, the size of the image to be processed will be very small, and the method of extracting local features from the solution will not work well. In this study, no steps have been taken during the resolution to allow a larger area to search for other constraints, and a buffer pool [6, 1].

The introduction of nonlinear function theory is mainly to improve the teaching ability of the network and to make deep neural multi- points.

If it is necessary to improve the parameters related to the nonlinear network, for example, the number of network layers, the GA can get a better effect, the effect of the modified PLOS is that the specified length is fixed, so it is not suitable for the dynamic swarm model [17].

3.3. Optimized process design. The flow chart of CNN optimization by using the improved chameleon swarm algorithm proposed in this paper is shown in Figure 3.1, where Y indicates that conditions are met, while N indicates that conditions are not met.

In this study, the learning algorithm of the algorithm neural network is calculated as a particle swarm algorithm to construct the chameleon swarm algorithm. Therefore, the number of process particle is the

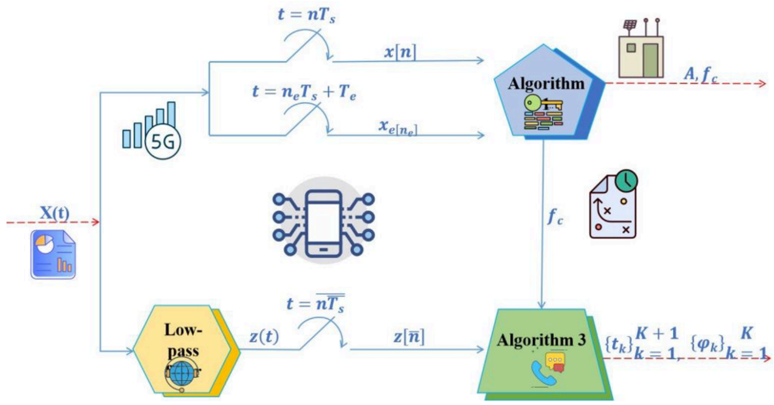

Fig. 3.1: Improved Chameleon Swarm Algorithm -CNN Training flowchart

number of network courses. First, start the task and speed of this product, then calculate the fitness function according to the error of actual result and requirement, and use the world look good and speed of each bit to get the weight of the network. Those. New weights are then substituted and iterated so that the algorithm stops until the fitness changes to some threshold [4, 15].

3.4. Optimization and improvement. Ideally, the performance of each sub-scale should be assessed in the same way as the last stage of training, the same duration, the same number of training sessions and so on. But this is not true, because if the number of particles is M and the number of epochs optimized by the chameleon swarm is N, then the optimum time of this parameter is MN. But there are too many optimization parameters in ANN, which results in a long training time and a high cost.In addition, many regular training in the database can lead to excessive interference, resulting in a number of performance issues in practical applications.It should be noted that during the optimization process, there is no need to know a small proportion of the values of each parameter for optimal performance. Only the value of the parameter is superior to other parameters, which means that its function is the best solution for particle operation. Therefore, this article proposes two development methods to improve efficiency. When CNN control realizes capability, it reduces the time parameter optimization and shortens the training time.

Dependency classification model: Parameter resolution is dependent on data size and quality. That is to say, when interpreting the data, we can estimate the number of iterations required to build an optimum neural network. As mentioned above, this article first introduces CNN network frequency, then selects the object as focus, and finally calculates Spearman correlation coefficient of particles according to the calculation results.The formula is as follows: (3.3) (3.4):

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(d_i)^2}{n_0} \tag{3.3}$$

$$d_i = \sigma(p_i) - \sigma(q_i), n_0 = n^3 - n \tag{3.4}$$

Based on the correlation coefficient, the predetermined epoch number threshold value E can be obtained, which can be used as the training turn value of CNN optimization by using the improved chameleon swarm algorithm [20, 8].

Process-based transformation: Although the results obtained from the correlation level based method are reliable, this process requires extensive training of CNN, which takes a lot of manpower resources. Therefore, this article presents another change based approach. Because of the randomness of backpropagation during initial training phase, the performance accuracy of the network structure is unstable. However, as training increases, the number of training cycles on data gradually increases, and the recognition accuracy often results

Table 4.1: The amount of data in the data set used for the experiment

	Train	Test	Val
CIFAR10	50000	10000	N/A
CIFAR100	50000	10000	N/A
Subset10	12081	1500	500
Subset20	37476	4500	1500
Subset50	59907	7500	2500

Table 4.2: Parameters to be optimized and their ranges

layer	hyper-parameter	dynamic range
Convolutional Layer	Number of feature maps	50∼180
	Padding size	0∼7
	Convolution kernel size	2∼7
Pooling Layer	Pooling type	MAX, AVE
	Pool core size	2∼7

in more stable performance, appearing in network structure Therefore, the stability of the network structure can be expressed as follows, shown in formula (3.5):

$$CV = \frac{\mu}{\sigma} \tag{3.5}$$

Here is the formula (3.6) (3.7):

$$\mu = \frac{\sum_i^N \text{accuracy}^k[i]}{N} \tag{3.6}$$

$$\sigma = \sqrt{\frac{\sum_i^N \left(\text{accuracy}^k[k] - \mu\right)^2}{N}} \tag{3.7}$$

$\text{accuracy}^k[k]$ is the classification accuracy of the i th particle in the k stage. Allowing comparison of different performance of the network structure in different cycles. Therefore, when the electric field is unstable, the best accuracy for the particle size can be obtained by comparison. This method will not require an optimal algorithm for the final classification, thus reducing the computation cost [3].

4. Interpretation of Result.

4.1. Experimental design. The data used in this article includes CIFAR10, CIFAR100, and ImageNet data, which can be divided into 10 classes, 20 classes, and 50 classes. respectively. Table 4.1 shows how many images are used for training and testing across all documents. In this paper, we apply Alexnet Network Model and Improved Chameleon Swarm Algorithm to optimize the data classification efficiency. In this paper, we use the simplified model of Alexnet to improve the precision of classification.

Parameters that need to be optimized in the training stage are shown in Table 4.2, and there are about 3.6×1020 possibilities for parameter setting. Therefore, even for the standard Alexnet model configuration, a simple and direct search is not possible. In this experiment, the hyperparameter of the improved chameleon swarm algorithm is set as c_{r1}= c_{r2}= 1.494,ω = 0.792.

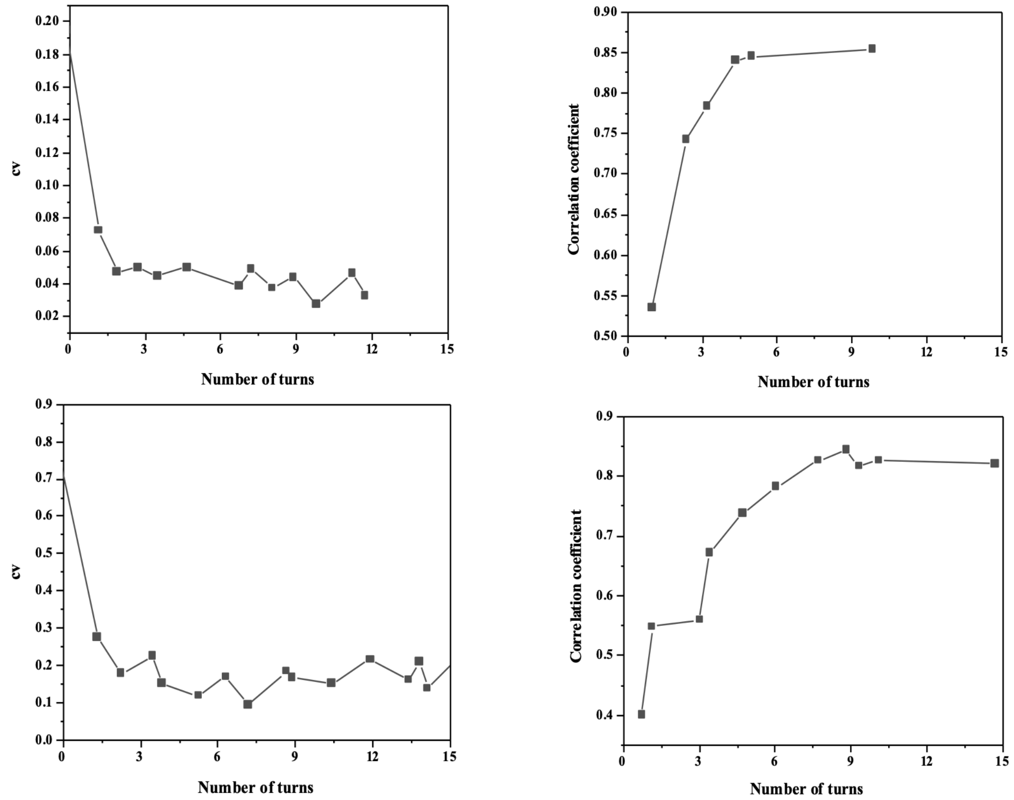

Fig. 4.1: Chart of variation of correlation coefficient between volatility and ranking

4.2. Analysis of experimental results. In this paper, the CNN model shown in Figure 4.1 (a) (b) (c) (d) is related to erêmbîn and the chameleon group algorithm is developed and optimized based on transformation. As can be seen from the figure, the relationship level will be above 0.8 and the change will gradually stabilize. However, evolutionary based technique does not require more quantitative analysis of data for comparison and better for optimization of neural network. As shown in Figure 4.1, in the next test, the number of training sequences for the CIFAR10 data and CIFAR100 data will be set to 5 and 10, respectively. The number of training sequences for the CIFAR10 data and CIFAR100 data will be set to 5 and 10, respectively. By limiting the number of training rounds, the load ratio of the improved chameleon swarm algorithm can be reduced as compared to the original improved chameleon swarm algorithm [2].

On the basis of improved chameleon clustering algorithm, this article studies the relationship between classification accuracy and iteration number of CNN model by parameter optimization. The number of optimization methods in this model is 21. Considering the training cost, the first number of bits per unit is 15, and the number of iterations is 0-60.Best and Average corresponds to the best and average of 15 sections. This figure confirms that the performance of the Alexnet network model improves as the iteration time increases. However, it should be noted that the development of the chameleon swarm algorithm for random initialization only approximates the network model to the optimal model and cannot guarantee that the best view of the world has been found [16, 7].

In Table 4.3, the image classification performance of Alexnet model optimization using improved chameleon swarm algorithm was compared with that of Alexnet model. In different data tested, the classification accuracy of network model was improved using improved chameleon swarm algorithm was superior to different level, but it was only 2% to 4% higher than the BP method usually used in the past.At the same time, it also shows that, because the improved chameleon swarm algorithm is essentially an optimization process based on random

Table 4.3: Improved Chameleon Swarm Algorithm -Alexnet performance compared to Standard Alexnet

Data set	Alexnet	Improved chameleon group algorithm-Alexnet	Performance difference	optimal cost
CIFAR10	77.76%	80.25%	2.48%	2%
CIFAR100	52.43%	55.67%	3.24%	3%
Subset10	72.57%	74.83%	2.26%	3%
Subset20	59.69%	65.46%	5.78%	4%
Subset50	57.56%	58.85%	1.29%	4%

allocation, no data set is necessarily the best, but the training results that are relatively close to the essential effect of the model can be found in a large number of training.Moreover, the performance of the model can be improved continuously by using the modified PGA [11, 19]. The convergence of the Chameleon algorithm is demonstrated. It is proved that the Chameleon Algorithm can converge to the global optimum with the increase of time.

5. Conclusion. This paper proposes a particle-particle optimization algorithm to optimize neural network constraints to solve neural network problems. Local minimum value due to multiple parameters. In this paper, the parameter setting and function selection of convolutional neural networks are explored and experimented, and the influence of these parameters on model training is revealed. The experimental results show that by pre-setting the relevant hyperparameters and matching with RMSProp or Adam, the accuracy of convergence can be reached faster, thus improving the training efficiency of convolutional neural networks.In the course of design, this article discusses how to reduce the number of computations by optimizing the data and controlling the number of training courses, in order to make the experience more satisfactory. This article tries to prove that by improving the chameleon swarm algorithm, the image recognition accuracy of the improved Alexnet model is 1.3 to 5.7 times higher than that of the traditional online training model. Meanwhile, the model proposed in this article is independent of the specific structure of Alexnet network model, so this model is universal and can be applied to most neural networks.

REFERENCES

[1] A. Ali, Y. Zhu, and M. Zakarya, *Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction*, Neural Networks, 145 (2022), pp. 233–247.

[2] A. Arabali, M. Khajehzadeh, S. Keawsawasvong, A. H. Mohammed, and B. Khan, *An adaptive tunicate swarm algorithm for optimization of shallow foundation*, IEEE Access, 10 (2022), pp. 39204–39219.

[3] M. S. Braik, M. A. Awadallah, M. A. Al-Betar, A. I. Hammouri, and R. A. Zitar, *A non-convex economic load dispatch problem using chameleon swarm algorithm with roulette wheel and levy flight methods*, Applied Intelligence, (2023), pp. 1–40.

[4] Y. Dong, Q. Liu, B. Du, and L. Zhang, *Weighted feature fusion of convolutional neural network and graph attention network for hyperspectral image classification*, IEEE Transactions on Image Processing, 31 (2022), pp. 1559–1572.

[5] M. Ghasemi, M.-A. Akbari, C. Jun, S. M. Bateni, M. Zare, A. Zahedi, H.-T. Pai, S. S. Band, M. Moslehpour, and K.-W. Chau, *Circulatory system based optimization (csbo): An expert multilevel biologically inspired meta-heuristic algorithm*, Engineering Applications of Computational Fluid Mechanics, 16 (2022), pp. 1483–1525.

[6] T. Ghazal, *Convolutional neural network based intelligent handwritten document recognition*, Computers, Materials & Continua, 70 (2022), pp. 4563–4581.

[7] G. Habib and S. Qureshi, *Optimization and acceleration of convolutional neural networks: A survey*, Journal of King Saud University-Computer and Information Sciences, 34 (2022), pp. 4244–4268.

[8] G. Hu, R. Yang, and G. Wei, *Hybrid chameleon swarm algorithm with multi-strategy: A case study of degree reduction for disk wang–ball curves*, Mathematics and Computers in Simulation, 206 (2023), pp. 709–769.

[9] S. Irene D, J. R. Beulah, A. K, and K. K, *An efficient covid-19 detection from ct images using ensemble support vector machine with ludo game-based swarm optimisation*, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 10 (2022), pp. 675–686.

[10] N. Jia, Y. Cheng, Y. Liu, and Y. Tian, *Intelligent fault diagnosis of rotating machines based on wavelet time-frequency diagram and optimized stacked denoising auto-encoder*, IEEE Sensors Journal, 22 (2022), pp. 17139–17150.

[11] X. Kan, Y. Fan, Z. Fang, L. Cao, N. N. Xiong, D. Yang, and X. Li, *A novel iot network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network*, Information Sciences, 568 (2021), pp. 147–162.

[12] G. Liu and W. Ma, *A quantum artificial neural network for stock closing price prediction*, Information Sciences, 598 (2022), pp. 75–85.

[13] M. Noroozi, H. Mohammadi, E. Efatinasab, A. Lashgari, M. Eslami, and B. Khan, *Golden search optimization algorithm*, IEEE Access, 10 (2022), pp. 37515–37532.

[14] C. B. Pop, T. Cioara, I. Anghel, M. Antal, V. R. Chifu, C. Antal, and I. Salomie, *Review of bio-inspired optimization applications in renewable-powered smart grids: Emerging population-based metaheuristics*, Energy Reports, 8 (2022), pp. 11769–11798.

[15] R. M. Rizk-Allah, M. A. El-Hameed, and A. A. El-Fergany, *Model parameters extraction of solid oxide fuel cells based on semi-empirical and memory-based chameleon swarm algorithm*, International Journal of Energy Research, 45 (2021), pp. 21435–21450.

[16] A. Sridharan, *Chameleon swarm optimisation with machine learning based sentiment analysis on sarcasmdetection and classification model*, Int Res J Eng Technol, 8 (2021), pp. 821–828.

[17] C. Tian, Y. Yuan, S. Zhang, C.-W. Lin, W. Zuo, and D. Zhang, *Image super-resolution with an enhanced group convolutional neural network*, Neural Networks, 153 (2022), pp. 373–385.

[18] A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf, *A review on modern defect detection models using dcnns–deep convolutional neural networks*, Journal of Advanced Research, 35 (2022), pp. 33–48.

[19] Y. Wang, X. Qiao, and G.-G. Wang, *Architecture evolution of convolutional neural network using monarch butterfly optimization*, Journal of Ambient Intelligence and Humanized Computing, 14 (2023), pp. 12257–12271.

[20] J. Zhou and Z. Xu, *Optimal sizing design and integrated cost-benefit assessment of stand-alone microgrid system with different energy storage employing chameleon swarm algorithm: A rural case in northeast china*, Renewable Energy, 202 (2023), pp. 1110–1137.

Edited by: B. Nagaraj M.E.
Special issue on: Deep Learning-Based Advanced Research Trends in Scalable Computing
Received: Aug 3, 2023
Accepted: Nov 3, 2023