Convolutional Neural Network Hyper-Parameter Optimization Using Particle Swarm Optimization

Md. Ferdouse Ahmed Foysal, Nishat Sultana, Tanzina Afroz Rimi, and Majedul Haque Rifat

Abstract CNN has recently gained popularity in the field of image processing. It has proven its niche in the field of machine learning. Computational models which use biological computation have been seen to use CNN. In this paper, we are going to optimize mainly one of these CNN hyper-parameters called convolution size. The optimization of parameter is actually the selection of the parameter by which the model will increase its performance. We are going to use evolutionary algorithm techniques. One of this techniques is called as particle swarm optimization(PSO). It is a community ground methods which select its population based on their fitness of the members where at the final selection the least fit members tend to stay at the population. We have various accuracy level by using only CNN and using CNN and PSO together.

Keywords CNN · PSO · Hyper-Parameter · Population

1 Introduction

CNN's level of success nowadays has got exceptional proportioned output. Back propagation in deep neural network for learning DNN has also got popularity. They proved their niche of work and justified it with proper application. The technolo-

Md. F. A. Foysal (✉) · T. A. Rimi · M. H. Rifat
Daffodil International University, Dhaka, Bangladesh
e-mail: ferdouse15-5274@diu.edu.bd

T. A. Rimi
e-mail: tanzinaafroz.cse@diu.edu.bd

M. H. Rifat
e-mail: majedul15-6784@diu.edu.bd

N. Sultana
Bangladesh University of Engineering and Technology,
ECE Building, Azimpur Road, Dhaka 1205, Bangladesh
e-mail: nishatsultana241@gmail.com

A. E. Hassanien et al. (eds.), *Emerging Technologies in Data Mining and Information Security*, Advances in Intelligent Systems and Computing 1300,
https://doi.org/10.1007/978-981-33-4367-2_35

gies are outperforming human resource to some extent. Parameters are an important concept related to them. The neural networks depend on their parameter can also termed as hyper-parameter. The tuning of these parameters is essential for expert practice. Automatic methods need to be designed. The models are needed to identify hyper-parameters. They are significantly necessary for compound DNN infrastructure. For this criteria, trial and error is speculative [1]. CNN belongs to the class of neural networks. It is a type of artificial neural network. This neural network is specifically designed for image processing. It is also used for image recognition. But it is dedicated designed for processing pixel data. Convolutional neural network got a special structure called neurons. They are like frontal lobe. This frontal lobe is used for the processing of visual stimuli. The competence of humans and machines is being divided by the introduction of artificial intelligence. Computer vision is a promising domain regarding CNN. Empowering machines is the main goal of the field. A nominal number of information's has been utilized. Image analysis, image and video recognition, media entertainment, suggestion systems, natural language processing can be taken as examples. Deep learning with computer vision has been built and nurtured with time. They are done with the help of CNN. Hyper-parameters can be of different types. Number of hidden units, learning rate are included in that. The parameters are set before the data has to be trained. Particle swarm optimization (PSO) is a meta-heuristic search algorithm. It got its inspirational movement from birds. Not birds actually the flocks of birds. It is widely used and accepted. It can be used to solve a wide range of problems. It mainly targets the optimization problems. PSO can be used in signal processing. It can be also used in graphics, robotics. They are diverse scientific fields. It is an optimization tool. The civil engineering arena has measured the success of the search technique. It has proved its niche in structural blueprint, structural state evaluation. It has also shown its position in health monitoring. It also has a handsome grip on structural material portrait and figuring. Shipment chain architecture is an heavy part of this. Traffic stream prediction, traffic mastery, traffic mishap prediction are its real-time applications. River level forecasting, structure upon of water/wastewater circulation chain plays an important role.

2 Associated Literature

In this paper, we are going to optimize single parameter of CNN. The technique we are going to use is PSO. PSO stands for particle swarm optimization. PSO is introduced by observing the nature of birds flock. In the past, deep neural network's hyper-parameters have been optimized using the PSO structure. PSO is a meta-heuristics algorithm. By using this algorithm, we are going to update only one parameter of CNN. Here, we have used digit classification dataset, where using only the CNN architecture gives accuracy of 94%. If we use CNN and PSO together, then we get the accuracy of about 95%. In previous works, PSO has been used to optimize the DNN's hyper-parameter. They have used the MINIST and occasionally CIFAR 10 dataset to automate the selection of the dataset [2].

3 Methodologies

Applying PSO to select hyper-parameter which will give the best result is a niche topic in the industry. It could imprint a valuable foot mark in optimizing the meta-heuristics algorithms family. PSO can be used to identify the hyper-parameter of DNN as well as CNN [2]. We are taking the DNN's hyper-parameter selection as the resource of this paper. The architectures that have been used are introduced in the consequent section of the paper.

3.1 Deep Neural Network

Deep neural network is hot cake in the recent research industry. It has proved its niche by its implementation and wide spreaders. From object detection to speech recognition, all the research fields are having DNN's contribution. Deep neural network is a kind of neural network [3]. It contains a certain level of complexity. It got more that two layers. Mathematical models are being used to process data. It can be used to detect pedestrians which will reduce the risk of road accident. It needs a large amount of data to perform well.

3.2 Convolutional Neural Network

Convolutional neural network is a kind of neural network to be used efficiently to identify image. It is mainly used for image processing. CNN has one or more convolutional layers. The layers are mainly used for image reciprocation. Classification of images and segmentation of images also use CNN. For autocorrelated data, CNN is used. A convolution is essentially sliding a filter over the input. CNN layers are not fully connected. To process pixel data of any image, CNN is artificial neural network which is being designed.

3.3 Hyper-Parameters

Hyper-parameters are variables that we need to set before applying a learning algorithm to a dataset. Hyper-parameters consist of two classes optimizer and model-specific parameters. Optimizer parameters are used for optimization and training process. They model specific parameter is used for determining the structure of the model [3]. Number of epoch, learning rate, minibatch size are some hyper-parameters which are known as optimizer parameter. Number of hidden nodes or units are defined as the model hyper-parameters.

3.3.1 Hyper-Parameter Selection

Optimal values for hyper-parameters selection have an impact of convolutional layers. Network depth, number of filters and their sizes are determined as the hyper-parameters. They have a drastic effect on the functionality of the classifier. Hyper-parameters can be selected in various ways. Automated hyper-parameter selection is based on two types of selection namely model-based and model-free selection [3]. If we consider the model-free selection, then we can have two variations of the selection types. It can function on grid search and random search. When we intend to use the model-based selection, we have various ways to select parameter. In this paper, we are going to use the model-based selection technique of evolutionary algorithms [4]. PSO is a evolutionary algorithm.

3.3.2 Model-Free Hyper-Parameter Selection

Grid and random search are two types of the model-based selection. It is commonly practise as the upon the DNN hyper-parameters. It performs better if the parameter quantity are low. A domain of values has to be choosen first. Then, exploration within this value continues. The vice versa technique is random search. This one is comparatively faster than the grid one. It cannot be dynamically updated during the experiment.

3.3.3 Model-Based Hyper-Parameter Selection

Model-based selection works on various techniques. Probabilistic methods are quite often used. Bayesian optimization techniques are used. It can have two types of variation in selection TPE and spearmint. Spearmint is being used for the hyper-parameter selection in DNN. RBF surrogate model is being used as the non-probabilistic method. In this paper, we are going to use the PSO technique which is a evolutionary algorithm.

3.3.4 Evolutionary Algorithms for Hyper-Parameter Selection

In artificial intelligence, evolutionary algorithm plays an important role. It is a component of evolutionary computation in the field of AI. Evolutionary algorithm (EA) works through a selection methodology. The population is being sorted according to their fitness. The member of population which is less fit is concluded to be eliminated. It could be a single of set of member. The fit members are allowed to survive [3]. They survive until the next good solutions are selected.

3.4 Particle Swarm Optimization

Swarm intelligence family holds a special algorithm called PSO. PSO is a population-based meta-heuristic technique. It selects a feasible solution first denoting as particles then changes its position by the evolution technique [5]. It can function on an large set of population holding many dimensions. It is simple yet efficient technique.

4 Benefaction

In this paper, we are going to introduce evolutionary algorithm to optimize the selection of the hyper-parameters of CNN. CNN has various hyper-parameters which plays an vital role in the result of the classifier. In this section, we are going to optimize one parameter [6]. Our optimized parameter is called as convolution size. Evolutionary algorithm is a biology-influenced technique for the selection of hyper-parameter. It is easy to correlate, and it is independent from target CNN.

5 Experimental Ground

5.1 Experimental Setup

Our hyper-parameter selection methodology was implemented on Python. We have used the NumPy library. Our CNN was trained using the Keras. In classical machine learning techniques, self-hyper-parameter selection does exists. Regularization of the regression is a good scope of optimization. Stochastic gradient descent optimization can be another fruitful chamber in this area. In all the experiments, the characterization of PSO was constraint.

5.2 Datasets

In this paper, we have given out our attention on the optimized selection of the parameter. We have used our synthetic datasets. We have not used any of the benchmark datasets [7]. Some of the ground breaking datasets are MNIST, CIFAR-10/100.

5.3 Experimental CNN Architecture

In this experiment, we are using an synthetic dataset which roughly contains ten 10,000 color images. They got 1000 images per class to be justified. Designing an CNN architecture based on the dataset was challenging. In this paper, we have implemented the typical CNN architecture. This consists of a convolutional layer. It also got max-pooling layer. Last of all, it got a fully connected layer. ConvNet is a class of DNN. In each layer, neurons are connected. They are connected with the next layer. This concept is called multilayer perceptions (Fig. 1).

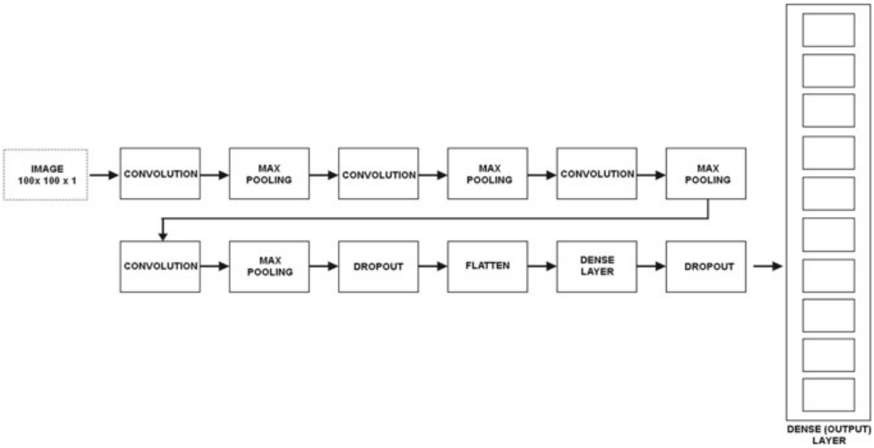


Fig. 1 Experimental CNN architecture

In our typical CNN model, we have in total of 13 layers. There are some layers defined as convolutional layers. In this architecture, we have three layers. Each class probability is defined as the softmax activation. There is a layer in the model defined as the flatten layer. In the last phase, we have two dense layers. The dropout layer parameter is 0.20. There are three max-pooling layers [8].

6 Analysis of Results

The experimental study what we have represented here is mainly partitioned into two sub-experiments. First of all, we have experimented the CNN. Using only CNN, the digit classification dataset has given us accuracy all about 94%. Whereas the PSO+CNN gives us accuracy all about 95%. At first, we have assumed that how CNN affects the hyper-parameter by using their own convolution size. Here, we have used CNN's typical structure. None of the VGG, Res or Dense net has been used. We have used the layer-based CNN where three different types of layers are present. Training data detects the certainty of the classifier model. Training accuracy is the model accuracy. Validation accuracy is defined differently. The figure below shows a graph which shows the accuracy of our model.

Training accuracy is defined as the situation when the classifier model is applied. The model is applied on training data. Validation accuracy is defined as the accuracy when the classifier model is applied on a few selected unknown data (Figs. 2 and 3).

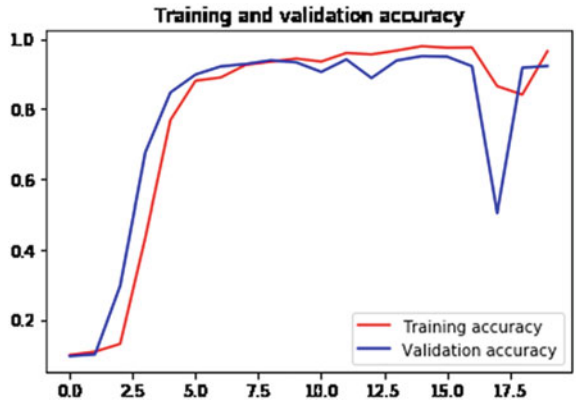


Fig. 2 Training and validation accuracy while applying CNN

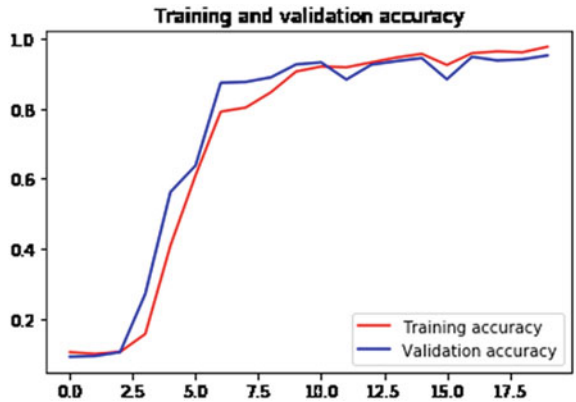


Fig. 3 Training and validation accuracy while applying PSO-CNN

Training loss got a different definition. The loss is actually occurred while training the dataset. Validation loss is stated as the trained network experiment on the data (Figs. 4 and 5).

Fig. 4 Training and validation loss while applying CNN

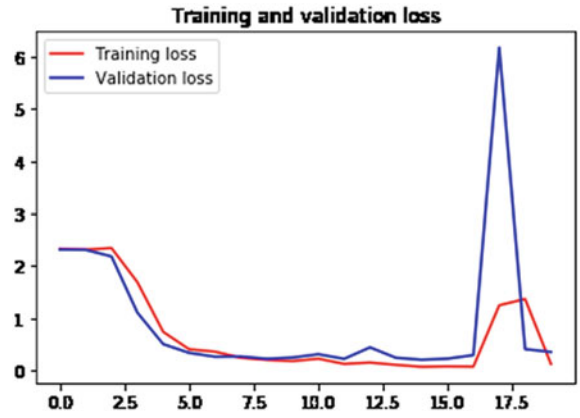
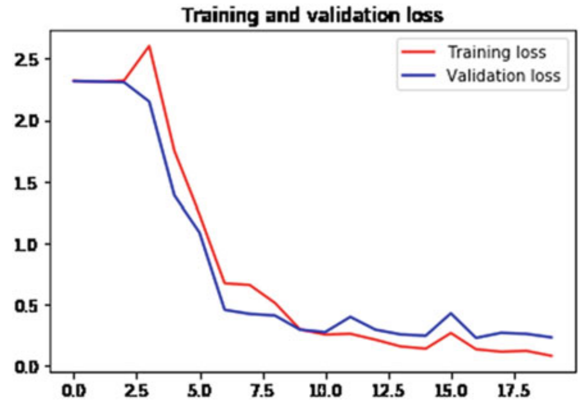


Fig. 5 Training and validation loss while applying PSO-CNN



7 Discussion of Results

PSO-CNN have acquired a decent accuracy of 95%, while CNN have acquired 94% accuracy. We have calculated the precision, recall and F1-score from the test dataset containing 2000 images.

The classification table is provided below (Table 1).

The classification table using the PSO and CNN both together is also given below (Table 2).

Now, the confusion matrix is in consideration. The confusion matrices of CNN and PSO-CNN are given below (Figs. 6 and 7).

Table 1 CNN classification result

Class	Precision	Recall	F1-score
0	0.94	1.00	0.97
1	0.95	0.82	0.88
2	0.99	0.96	0.97
3	0.98	0.93	0.95
4	0.94	0.99	0.97
5	0.91	0.97	0.94
6	0.97	0.96	0.97
7	0.97	0.99	0.98
8	0.98	0.94	0.96
9	0.87	0.92	0.89
AVG	0.95	0.95	0.95

Table 2 PSO-CNN classification result

Class	Precision	Recall	F1-score
0	0.91	0.99	0.95
1	0.95	0.90	0.92
2	0.97	0.98	0.98
3	0.98	0.95	0.97
4	0.92	0.99	0.95
5	0.93	0.99	0.96
6	0.98	0.96	0.97
7	0.98	0.96	0.97
8	0.98	0.92	0.92
9	0.94	0.90	0.92
AVG	0.96	0.95	0.95

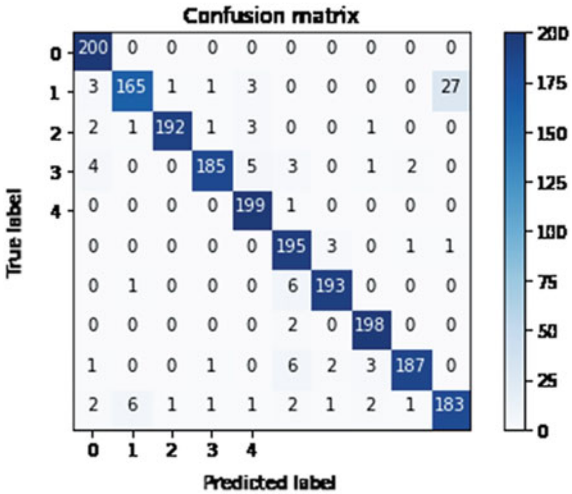


Fig. 6 CNN confusion matrix

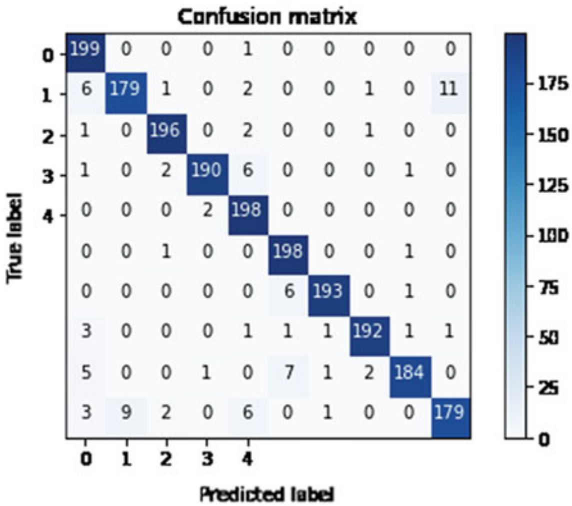


Fig. 7 PSO-CNN confusion matrix

8 Future Work

In future work, we will focus on dense layer size selection by giving preference on number of layers, number of neurons in every layer and the type of layer. Also, we will work on optimization function selection such as sigmoid function or than function for better performance.

9 Conclusion

In this paper, we present a modified form of convolution neural network by optimizing hyper-A better impressive performance is showed where we optimize convolution size in our model. Convolution neural network's hyper-parameters are optimized by using particle swarm optimization. The accuracy of our model for CNN architecture is 94, and if we use CNN PSO classification, then accuracy is about 95 after the optimization of CNN's parameter.

References

1. Loussaief, S., Abdelkrim, A.: Convolutional neural network hyper-parameters optimization based on genetic algorithms. Int. J. Adv. Comput. Sci. Appl. **9**(10), 252–266 (2018)
2. Soon, F.C., Khaw, H.Y., Chuah, J.H., Kanesan, J.: Hyper-parameters optimisation of deep CNN architecture for vehicle logo recognition. IET Int. Trans. Syst. **12**(8), 939–946 (2018)
3. Lorenzo, P.R., Nalepa, J., Kawulok, M., Ramos, L.S., Pastor, J.R.: Particle swarm optimization for hyper-parameter selection in deep neural networks. In: ACM, Proceedings of the Genetic and Evolutionary Computation Conference, New York, USA, pp. 481–488 (2017)
4. Foysal, M.F.A., Islam, M.S., Karim, A., Neehal, N.: Shot-net: a convolutional neural network for classifying different cricket shots. In: International Conference on Recent Trends in Image Processing and Pattern Recognition, pp. 111–120 (2018)
5. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: towards real-time object detection with region proposal networks. Adv. Neural Inf. Proc. Syst. 91–99 (2015)
6. Lee, W.-Y., Ko, K.-E., Geem, Z.-W., Sim, K.-B.: Method that determining the Hyperparameter of CNN using HS algorithm. J. Korean Inst. Int. Syst. **27**(1), 22–28 (2017)
7. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2961–2969 (2017)
8. Ye, F.: Particle swarm optimization-based automatic parameter selection for deep neural networks and its applications in large-scale and high-dimensional data. PLoS ONE **12**(12), e0188746(2017)