Swarm Intelligence Algorithms for Convolutional Neural Networks

Eva Tuba[1 [0000−0003−4866−9048]] and Ira Tuba[1 [0000−0002−2863−7084]]

Faculty of Informatics and Computing, Singidunum University, Belgrade, Serbia
etuba@ieee.org, ituba@ieee.org

Abstract. Digital images made a big difference in everyday life and science. They became crucial part in various areas such as medicine, astronomy, agriculture, etc. A common problem in applications with digital images is classification. Since recently, convolutional neural network has been widely used for image classification and classification accuracy is dramatically improved. To have the best possible classification accuracy of a CNN for a specific application, the necessary step is tuning various hyperparameters such as the number of different layers, number of neurons in each layer, optimization algorithm, activation functions, kernel size, optimization algorithm, etc. Recent studies show that swarm intelligence algorithms used for CNN's hyperparameters tuning give promising results and further research is needed. In this paper, a short analysis of swarm intelligence algorithms and their application to tuning CNN hyperparameters is presented.

Keywords: swarm intelligence · convolutional neural networks · deep statistical analysis.

1 Introduction

A great scientific progress has been made in the last several decades due to digital images. By replacing analog images with digital ones, various scientific branches were able to tremendously improve the quality of the results obtained by analyzing images while images began to be used only after the digital images have been introduced. For example, images have been used in medicine for hundreds of years, but with the usage of computer power and digital image analysis, the quality and the speed of the work have been increased drastically. Further, images were not useful in agriculture before, but nowadays, it is possible to analyze digital images obtained from air, by drone, or even satellite images of crop fields and detect plant diseases, soil quality and more. The application of digital images is very wide and appropriate processing methods were an interest of researchers for decades. Several years ago, convolutional neural networks (CNN) were used for image classification task and outperformed all previous methods by a large percent. Before that, with a slight adjustment of the existing classifiers, the results could be just a little bit improved. The huge potential of CNN attracted the attention of the researchers and nowadays there are numerous applications of the CNN on various problems [4, 9, 11].

After the initial radical improvement with CNN, further improvements in classification accuracy are possible by tuning CNN hyperparameters. Each image classification problem is specific up to a certain point. The quality of the classification by CNN differentiate depending on numerous hyperparameters that determine the architecture and mathematical model of the CNN. These hyperparameters include a number of different layers, order of layers, kernel size in each layer, loss function, optimizer and many more. Hyperparameters can be integers or real numbers and for each hyperparameter the set of allowed values is infinite. Therefore, CNN hyperparameter tuning represents a hard optimization problem. In the previous years, this topic was widely studied and different methods for CNN tuning were proposed. One method of finding the optimal values for hyperparameters is to use evolutionary or swarm intelligence algorithms [1, 8, 2].

Nowadays, there are numerous swarm intelligence (SI) algorithms and choosing the right one can be challenging so the first step in our research was to test the quality of SI algorithms and then apply the best one for tuning CNN hyperparameters. Initial research includes a small subset of hyperparameters that are tuned by the SI algorithm.

The rest of the paper is organized in the following way. Section 2 presents short description of the SI algorithms and deep statistical comparison method used for ranking algorithms. Section 3 describes the convolutional neural network and application of SI algorithm to hyperparameter tuning. Preliminary results of this research are presented in Section 4. Conclusion and future research plans are presented in Section 5.

2 Swarm Intelligence Algorithms

Swarm intelligence algorithms have been used for solving hard optimization problems for decades. The particle swarm optimization [14] and the colony optimization [7] are among the first algorithms in this group. Since the quality of the solutions generated by the SI algorithms were rather good, a search for better and improved algorithms has become a hot research topic. Nowadays, the number of SI algorithms grow tremendously and it is hard to tell which algorithm is better than another for certain problems. Moreover, the inspiration from nature and real life situations was endless which resulted in numerous algorithms that have similar or the same mathematical base but a different story where the equations are coming from. At the core, all SI algorithms are iterative optimization algorithms that have exploitation and exploration ability. A Swarm of simple agents performs a simple operation and exchanges information that guides their next movement. Collectively, after a certain number of iteration, the optimal or near optimal solution is found. The difference in SI algorithms is in the exploration and exploitation operators and the balance between them.

In order to choose the appropriate SI algorithm for CNN hyperparameter tuning, we initially used deep statistical comparison [3] to determine which algorithm has better exploration and exploitation operators. Deep Statistical

Comparison (DSC) is used for statistical comparison of meta-heuristic stochastic optimization algorithms such as SI algorithms. The main idea is to analyze the quality of the algorithm based on the distributions of the solution obtained from multiple runs. Each algorithm is started multiple times and the distributions of the solutions were compared between themselves. Ranking algorithms by DSC is more robust in comparison to the single statistic like average or median. In our previous work [10], we compared five widely used SI algorithms, particle swarm optimization (PSO), artificial bee colony (ABC) [5], firefly algorithm (FA) [12], flower pollination algorithm (FPA) [13], and bare bones fireworks algorithm (BBFWA) [6]. The results showed that for small dimension problems there is no significant difference between all five algorithms regarding the obtained solution, but the further analysis showed that the FA has the best exploration ability. Based on these results, we used FA for CNN hyperparameter tuning.

3 Swarm Intelligence Algorithms applied to Convolutional Neural Networks

Convolutional neural networks (CNN) are type of deep neural networks that are suitable for classification problems where each input represents a set of correlated data such as voice signals or images. The CNN architecture consists of convolutional, pooling and fully connected layers. Each layer has its own set of hyperparameters that should be tuned. The number of hyperparameters is large and includes kernel size, padding, type of pooling layer, etc. Each of these hyperparameters is a variable in the optimization problem where the goal is to maximize the classification accuracy.

Application of SI algorithm, in this case, the FA, to CNN hyperparameter tuning is not straightforward. It is necessary to adjust the FA for application to integer variables and some fitness function tweaking since fitness function evaluation is a relatively computationally and time-demanding operation. Our initial research includes simple adjustment for integer variables where simple rounding was performed for each generated solution. In this first research, the fitness function was classification accuracy while further research would include adjustment of this model so less fitness function evaluations are required.

4 Simulation Results

The FA for CNN hyperparameters tuning was implemented in Python 3.8 with PyTorch library. The simulations with Intel Core i7-10700K CPU at 5GHz, 16GB RAM, NVIDIA RTX 2060 graphic card, and Windows 10 Professional OS.

MNIST, the standard benchmark dataset for the image classification problem was used. This dataset contains 70,000 pre-processed images of handwritten digits. Each image is of size 28x28 in grayscale. For training, we used 60,000 images (6,000 images of each digit) and 10,000 was used for testing created CNN. The results obtained by the proposed FA were compared to LeNet-5 network.

Hyperparameters that were searched by the FA were kernel size for each layer, the number of feature maps and padding for convolutional layers. Architecture is based on LeNet-5 presented in Fig. 1. This CNN consists of two convolutional layers, each one followed by the pooling layer and at the end of the network are two fully connected layers.

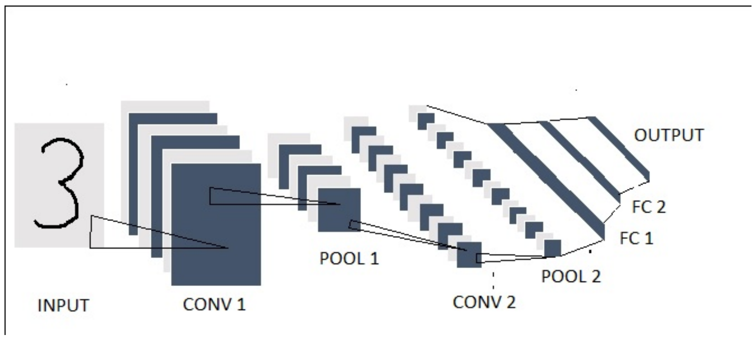

Fig. 1. Architecture of the used CNN

The parameters of the FA algorithm were set based on the previous experience and several test runs while the future work will include a deeper analysis of their influence. The parameters were set as follows: population size = 5, α=0.5, $\beta_0 = 0.2$, γ=1 and the maximal fitness function evaluation was 40.

Obtained classification accuracy is presented in Table 1. The accuracy presented in Table 1 for the network optimized by the FA (FA-CNN) represents the average accuracy obtained in 20 runs, i.e. the FA was started 20 times and the accuracy of the generated CNN (when testing on the unknown data, test set) were recorded. The standard deviation in the classification accuracy was 0.06. Based on the initial testing, it can be concluded that the FA can improve the CNN accuracy by tuning hyperparameters. Future research will include more hyperparameters such as optimizer used in CNN, dropout rate, learning rate, number of nodes in the fully connected layers, etc.

Table 1. Classification accuracy of LeNet-5 and FA optimized networks

CNN	Accuracy
LeNet-5	98.94
FA-CNN	99.16

5 Conclusion

Convolutional neural networks have been proved to be a superior method for digital image classification. In order to obtain the best possible results, CNN hyperparameters should be fine-tuned which is an exponential problem. The SI algorithms are an appropriate method for tackling this problem. In this paper, we tested the firefly algorithm and the preliminary results are encouraging. Future research will include better adjustment of the FA for CNN hyperparameters tuning. Besides including more hyperparameters, we will also test different architectures, which can also be a variable in the optimization problem.

References

1. Bacanin, N., Bezdan, T., Tuba, E., Strumberger, I., Tuba, M.: Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristics. Algorithms **13**(3), 67 (2020)
2. Bacanin, N., Tuba, E., Bezdan, T., Strumberger, I., Jovanovic, R., Tuba, M.: Dropout probability estimation in convolutional neural networks by the enhanced bat algorithm. In: 2020 International Joint Conference on Neural Networks (IJCNN). pp. 1–7. IEEE (2020)
3. Eftimov, T., Korošec, P., Seljak, B.K.: A novel approach to statistical comparison of meta-heuristic stochastic optimization algorithms using deep statistics. Information Sciences **417**, 186–215 (2017)
4. Jayanthi, J., Jayasankar, T., Krishnaraj, N., Prakash, N., Sagai Francis Britto, A., Vinoth Kumar, K.: An intelligent particle swarm optimization with convolutional neural network for diabetic retinopathy classification model. Journal of Medical Imaging and Health Informatics **11**(3), 803–809 (2021)
5. Karaboga, D., Basturk, B.: Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. Foundations of Fuzzy Logic and Soft Computing, LNCS **4529**, 789–798 (2007)
6. Li, J., Tan, Y.: The bare bones fireworks algorithm: A minimalist global optimizer. Applied Soft Computing **62**, 454–462 (2018)
7. M., D., L.M., G.: Ant colony system: a cooperative learning approach to the traveling salesman problem. IEEE Transactions on Evolutionary Computation **1**(1), 53–66 (1997)
8. Sinha, T., Haidar, A., Verma, B.: Particle swarm optimization based approach for finding optimal values of convolutional neural network parameters. In: Congress on Evolutionary Computation (CEC). pp. 1–6. IEEE (2018)
9. Tuba, E., Bačanin, N., Strumberger, I., Tuba, M.: Convolutional neural networks hyperparameters tuning. In: Artificial Intelligence: Theory and Applications, pp. 65–84. Springer (2021)
10. Tuba, E., Korošec, P., Eftimov, T.: In-depth insights into swarm intelligence algorithms performance. In: Modelling and Development of Intelligent Systems. pp. 334–346 (2021)
11. Vijh, S., Sharma, S., Gaurav, P.: Brain tumor segmentation using otsu embedded adaptive particle swarm optimization method and convolutional neural network. In: Data visualization and knowledge engineering, pp. 171–194. Springer (2020)
12. Yang, X.S.: Firefly algorithm, lévy flights and global optimization. Research and Development in Intelligent Systems **XXVI**, 209–218 (2010)

13. Yang, X.S.: Flower pollination algorithm for global optimization. In: International Conference on Unconventional Computing and Natural Computation. pp. 240–249. Springer (2012)
14. Zamfirescu, C.B., Filip, F.G.: Swarming models for facilitating collaborative decisions. International Journal of Computers Communications & Control **5**(1), 125–137 (2010)