Swarm Intelligence Algorithms for Convolutional Neural Networks

-Exploratory paper-

Eva Tuba, Ira Tuba

Singidunum University, Belgrade, Serbia

2nd Workshop on Evolutionary and Population-based Optimization (WEPO)

30th November 2021

Artificial Intelligence

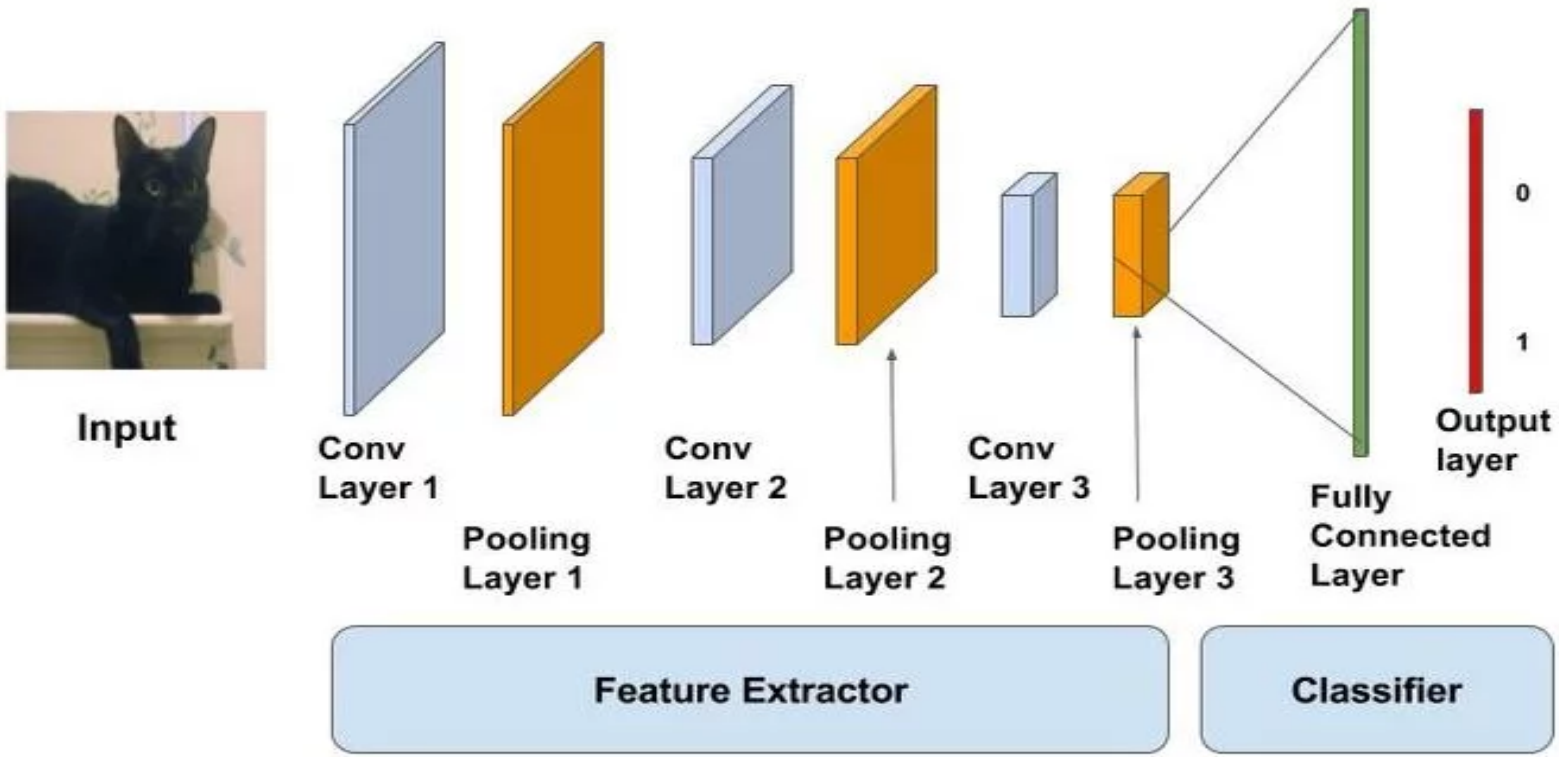
- Nowadays artificial intelligence has been successfully applied to numerous practical problems and these applications are used in everyday life:
  - autonomous vehicles,
  - medicine,
  - conversion voice to text,
  - object detection,
  - face recognition,
  - etc.
- Many of these problems have been solved by using deep learning methods and one of the recent deep learning method is convolutional neural network.

Convolutional Neural Networks

- Convolutional neural network (CNN) is a particular type of neural networks, that is broadly used in computer vision algorithms
  - for image classification,
  - image segmentation,
  - object detection,
  - face recognition,
  - etc.
- The difference between other classifiers and the CNNs is that CNN is applied to problems where inputs are not independent but rather correlated:
  - Digital images are represented as matrices of pixels where adjacent pixels within an image are correlated.
  - Time series data (e.g. voice, EEG) can be treated as static spatial data in CNN.
- CNNs are built from layers which operate in a sliding window fashion, i.e. convolution by kernel thus neighbor data are treated together.

Convolutional Neural Networks

- The classic CNN architecture consists of different types of layers, such as:
  - convolutional layer,
  - pooling layer and
  - dense layer (also called fully-connected or FC layer) at the end of the architecture



Convolutional Neural Networks

- Hyperparameter optimization for CNN architecture design optimizes the number of different hyperparameters, such as
  - convolutional layers,
  - filters,
  - selects the right filter size,
  - stride,
  - padding,
  - number of dense layers, along with the number of its neurons,
  - determines the learning rate,
  - epoch number,
  - and batch size.
- Also choosing the appropriate activation function, regularization technique to prevent overfitting and optimization algorithm for training the network is important for obtaining high classification accuracy.

Convolutional Neural Networks

- CNN architectures are handcrafted by using different techniques, such as grid search, random search and brute-force methods, that all require a lot of effort, domain expertise and time.
- Since, metaheuristics found to be effective in tackling NP-hard problems like this hyperparameter optimization problem could be solved by using some of the available metaheuristics.
- We propose an automatic approach for the selection of the appropriate hyperparameters by using a swarm intelligence approach.

Swarm Intelligence

Recently, swarm intelligence become prominent using the fact that extremely simple individuals exhibit miraculous collective intelligence. Examples include ant colonies, honey bees colonies, flocks of birds, schools of fish etc.

- These nature inspired metaheuristics simulate various natural phenomena. We talk about bee colony food finding or ant colony path finding, but in essence, in all these diverse mimicking we do two things.
- Nowadays, the number of SI algorithms grow tremendously and it is hard to tell which algorithm is better than another for certain problems.
- At the core, all SI algorithms are iterative optimization algorithms that have exploitation and exploration ability.
- The difference in SI algorithms is in the exploration and exploitation operators and the balance between them.

Deep Statistical Comparison

- Deep Statistical Comparison (DSC) is used for statistical comparison where the main idea is to analyze the quality of the algorithm based on the distributions of the solution obtained from multiple runs.
- In our previous work, we compared five widely used SI algorithms, particle swarm optimization (PSO), artificial bee colony (ABC), firefly algorithm (FA), flower pollination algorithm (FPA), and bare bones fireworks algorithm (BBFWA).
- 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.

Application of the SI to CNN

- Application of SI algorithm, in this case, the FA, to CNN hyper-parameter tuning is not straightforward.
- 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 for higher classification accuracy are required.

Simulation Results

- MNIST classification dataset consists of 28x28 pixel sized grayscale images of handwritten digits (0-9).
- Images are divided into 60,000 training and 10,000 test samples.

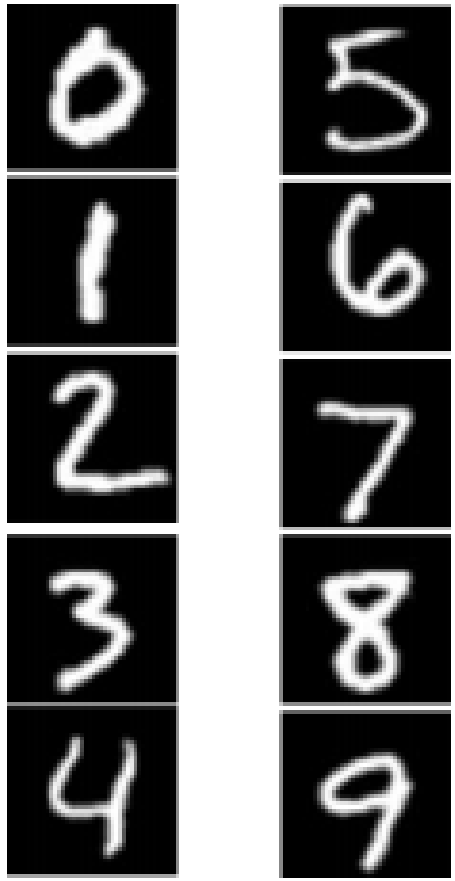


Figure: Example of images in MNIST dataset

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.
- The control parameters of the FA that are used in the proposed FA-CNN framework are:
  - population size $= 5$,
  - α=0.5,
  - $\beta_0 = 0.2$,
  - γ=1 and
  - the maximal fitness function evaluation was 40.

Simulation Results

- The proposed method is compared with the LeNet-5 network.

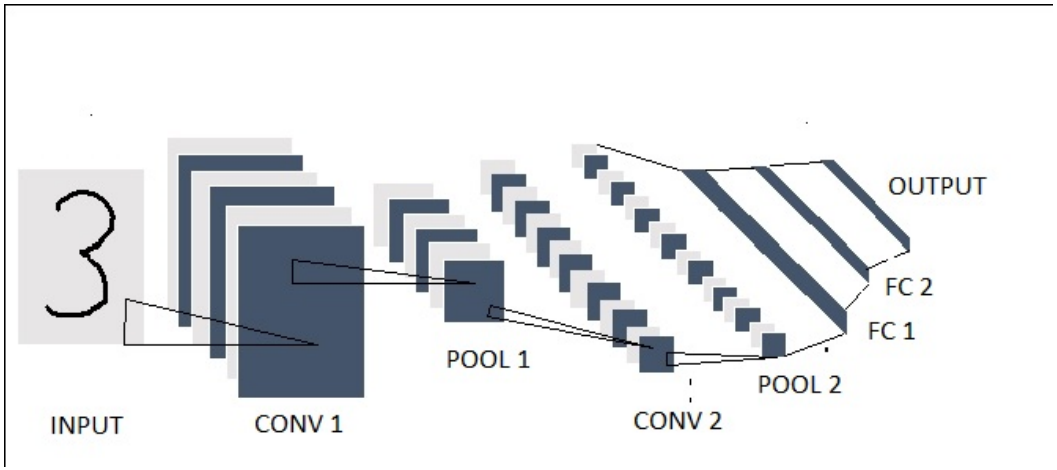


Figure: Architecture of the used CNN

- Hyperparameters that were considered in this research are:
  - kernel size for each layer,
  - the number of feature maps and
  - padding for convolutional layers.

Simulation Results

Table: Classification accuracy of LeNet-5 and FA optimized networks

CNN	Accuracy
LeNet-5	98.94
FA-CNN	99.16

Conclusion

- Convolutional neural networks have been proved to be a superior method for digital image classification.
- 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.

THANK YOU!