Name: Benedek Szücs, BSc student

Project type: thesis project

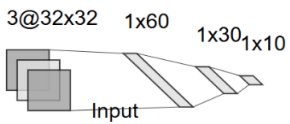
Topic: Neural Network structure optimalization for FPGAs

Supervisors: Dr. Balázs Bánhelyi

It is evident that Field Programmable Gate Arrays (FPGAs) are constrained in terms of processing power. Consequently, it is necessary to devise an algorithm to identify the optimal structures for specific tasks. The objective is to ascertain the technical details of the FPGA and to optimise a structure that can be used to train a neural network for the given task.

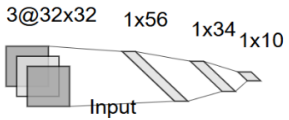
In our first experiment we tested two very similar networks on the famous CIFAR 10 dataset to see how much difference does a minimal change make

Network A



Attempt 1:	**50.84 %**
Attempt 2:	**49.60 %**
Attempt 3:	**49.79 %**
Average:	**50.07 %**

Network B

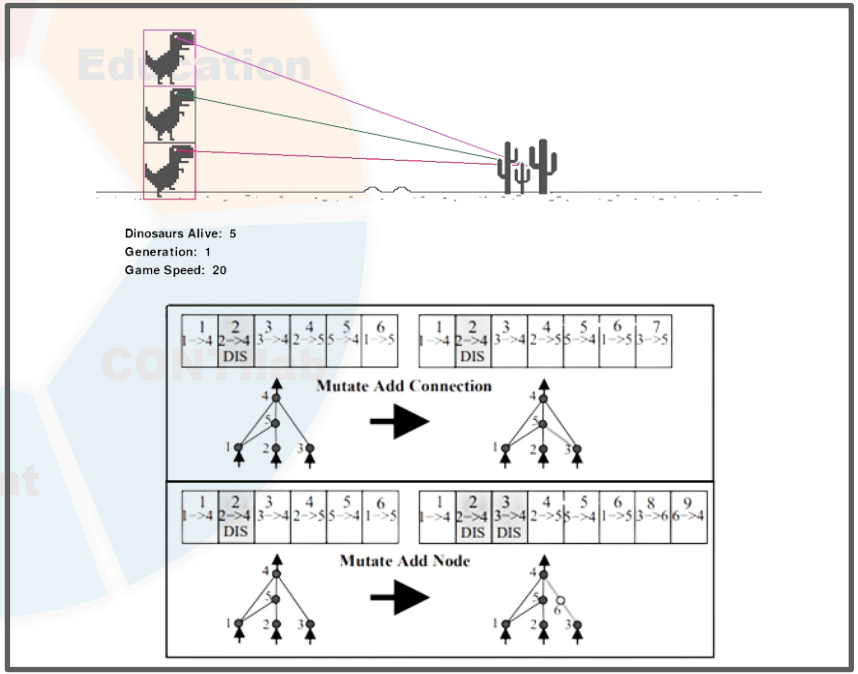


Attempt 1:	**51.08 %**
Attempt 2:	**50.60 %**
Attempt 3:	**50.35 %**
Average:	**50.67 %**

Experiments

Initially, the NEAT Neuroevolution approach was employed to assess the efficacy of diverse neural network architectures. This proved to be a fruitful endeavour, with the Google Chrome Dino game serving as a suitable benchmark. Subsequently, the NEAT algorithm was deployed on the MNIST dataset.
However, the resulting performance was suboptimal, with a notable lack of speed and the inability to achieve accurate number recognition. This was attributed to the inherent limitations of NEAT.

Consequently, we proceeded to utilise gradient optimisation on the mutated networks, hypothesising that this would facilitate more accurate estimation of the images. However, we encountered difficulties when attempting to translate the evolved networks into PyTorch networks, due to the limitations of PyTorch.



Results & future work

In the future, we intend to apply the principles of neuroevolution to fully connected PyTorch models.
We will investigate the use of generative algorithms to establish an objective function that takes FPGA parameters as input and outputs the most suitable neural network architecture for the given constraints and dataset.
This approach is designed to streamline the design process of neural networks tailored for FPGAs, enhancing their performance.