	1	Σ
Ibrahim Aabaza		
Marius Räther		

Assignment Sheet Nr. 5

Documentation

The network architecture used , contains 5 hidden layer of 64 neutrons each. The activation functions of these layer ReLU and Tanh in alternating fashion, starting with ReLU. A deeper or wider network would not give better results and the alternation of activation function did improve the results to some extend.

The network was trained with a learning rate of 0.05 which improved the results slightly. The network was trained for 200 episodes in total, although 100 episodes could already have been enough for roughly the same result.
The other hyperparamters and the optimizer were used as given with the original file.

The training dataset was extended to contain the horizontally and vertically flipped images as well.

The validation accuracy was:
88.91102257636122% And the network contains