Report of Exercise 02

Implementation Approach

This Exercise was approached following the lecture slides and the tensorflow tutorial on tensorflow.org.

Results

Changing the Learning Rate

In the following Figure 1 results with various learning rates are shown.
The used learning rates were: $\{0.1, 0.01, 0.001, 0.0001\}$. Every training session lasted 30000 epochs.

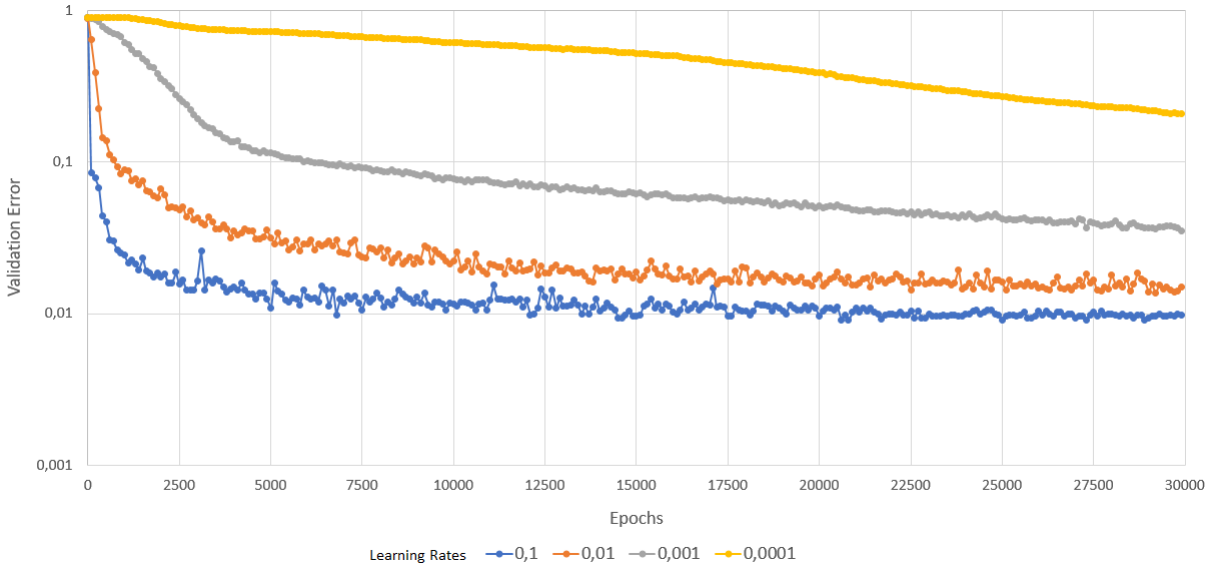

Figure 1: Different Learning Rates

It can be seen in Figure 1 above that while for 0.1 as a learning rate the validation error quickly converges toward around 1%, using the much smaller 0.0001 as a learning rate doesn't show any convergence after 30000 epochs and the validation error is still above 10%. In general it can be seen that for a larger learning rate the validation error will converge more quickly. This doesn't necessarily indicate that after more epochs of training the validation error for a small learning rate will not converge. But for this case the largest learning rate seems to be the best choice.

Runtime

The following Figure 2 shows results for a training session of 30000 epochs, run on a GPU. For this experiment the number of filters was varied. These number of filters were used: {8, 16, 32, 64, 128, 256}.
To avoid too extreme results, a learning rate of 0.01 was chosen for this experiment.

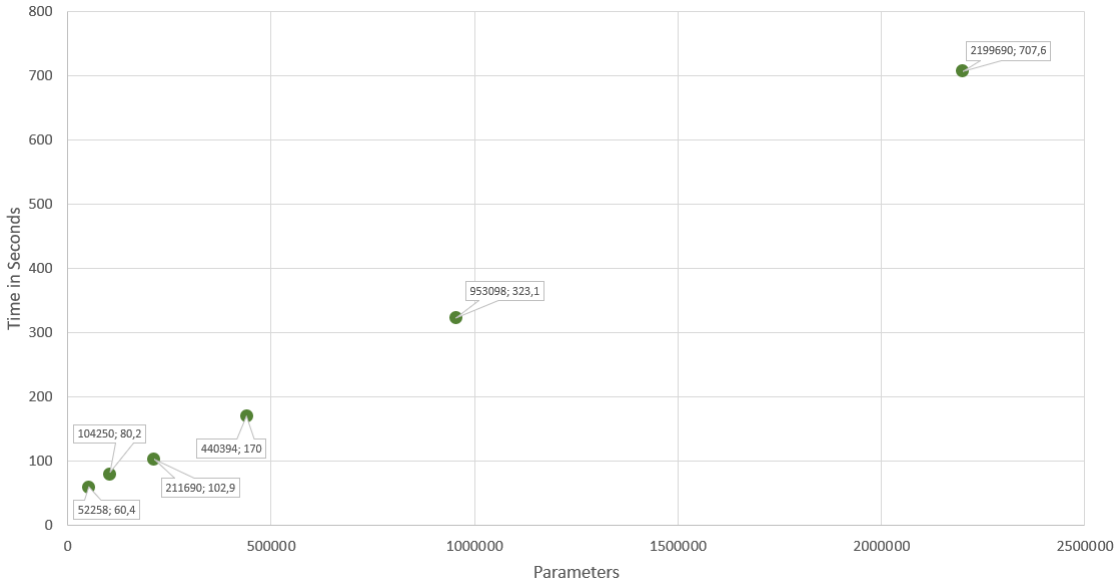

Figure 2: Different parameters, run on GPU

The number of parameters was calculated according to the lecture slides.
If the number of filters increases, the number of parameters as well as the runtime increases, too. The former conclusion is a result taken from calculating the number of parameters. The latter can be seen, transitively, in Figure 2.

In Figure 3 below the results for a training session of 30000 epochs, run on a CPU, are shown. In contrast to before, the number of filters used were {8, 16, 32, 64}. The learning rate was still 0.01.

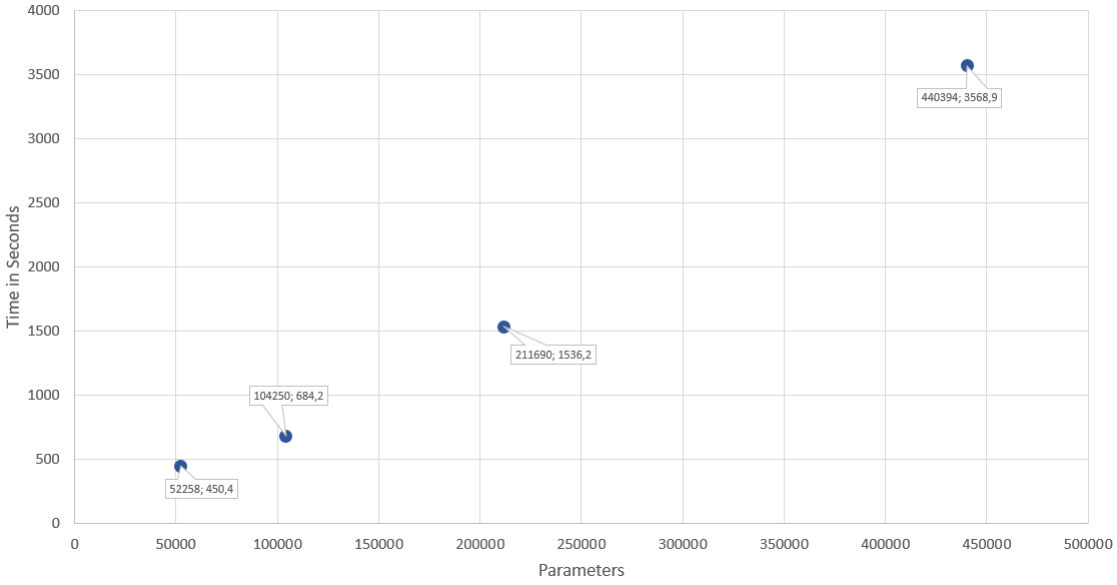

Figure 3: Different parameters, run on CPU

Comparing Figure 3 to Figure 2, clearly, a training session took a lot longer on a CPU then on a GPU. Even though the number of filters in the former experiment was higher, a training session still took significantly less time than when running it on a CPU.