Results

As seen in Figure 2 the results for the biggest learning rate 0.1 are best while for the smallest learning rate the Network converges very slowly. A to slow learning rate could lead to not converging at all while a to high learning rate could lead to divergence. The plots for different filters show the exact same picture of 0.1 being best and 0.0001 converging very slowly. The difference in filter size leads to a difference in the spatial observation of the filters. With a filter of size 1 there is no spatial information used. With a filter size of 7 the first layer will see an area of 7 by 7 and in the second layer each node has the information of an area of 14 by 14 from the input image. Bigger filters are thus better in scenarios where large patterns have to be matched. On the other hand smaller filters have less parameters and so the network with smaller filters is less likely to overfit. To see the difference between the filter sizes I plotted the different learning curves for one learning rate in one plot. In all cases we can see that the training error is lower the bigger the filters which is expected because of the higher degree of freedom. In Figure 5 we can see that the bigger filters handle the lowest learning rate better than the smaller filters.

The random search resulted in a best configuration of:

- batch_size = 23
- filter_size = 5
- num_filters = 44
- learning_rate = 0.0001913266

A higher learning rate of 0.1 and 10 epochs resulted in a test error of 0.0078. The best configuration learning rate uses warmstarts, which distorts the result.

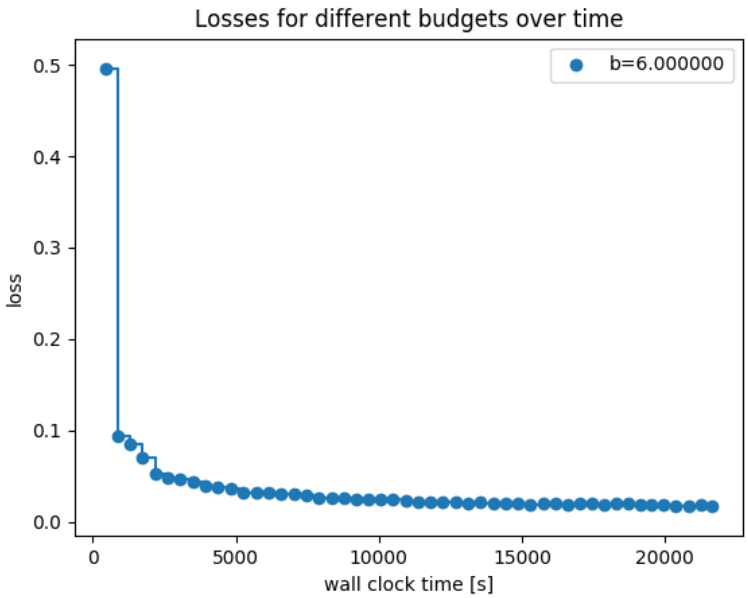

Figure 1: Loss of best configuration found by random search

```
lr=0.1    validation error 0.0138
lr=0.01   validation error 0.0351
lr=0.001  validation error 0.0831
lr=0.0001 validation error 0.7152
```

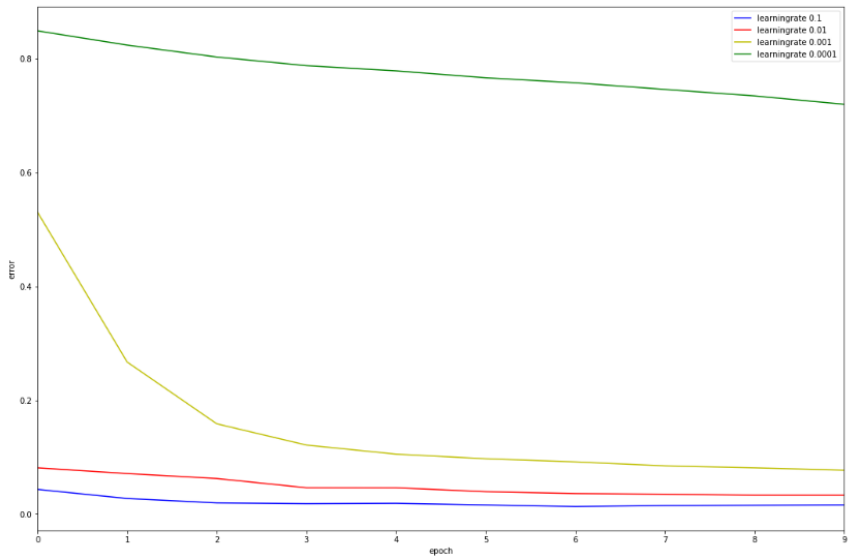

Figure 2: learning curve for different learning rates and filter size 3

```
filter_size 1    validation error 0.0539
filter_size 3    validation error 0.0351
filter_size 5    validation error 0.0232
filter_size 7    validation error 0.0249
```

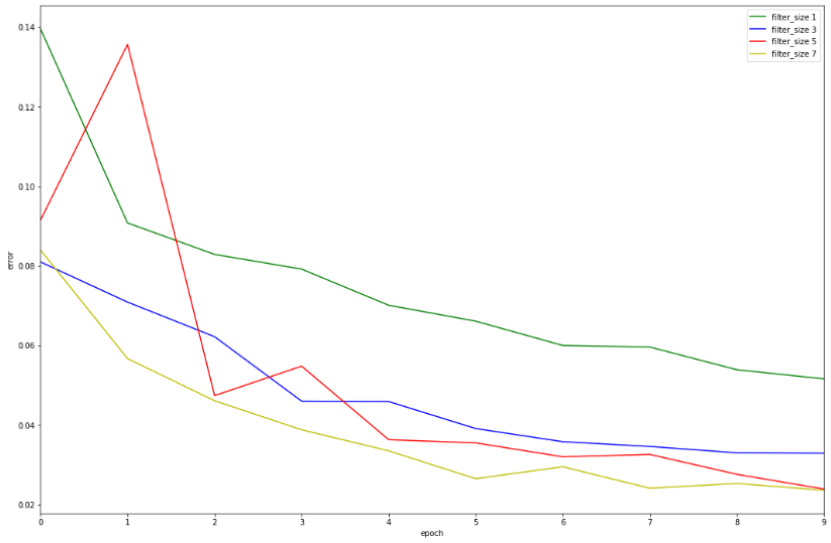

Figure 3: learning curve for different filter size and learning rate 0.01

```
filter_size 1    validation error 0.1211
filter_size 3    validation error 0.0831
filter_size 5    validation error 0.0817
filter_size 7    validation error 0.0738
```

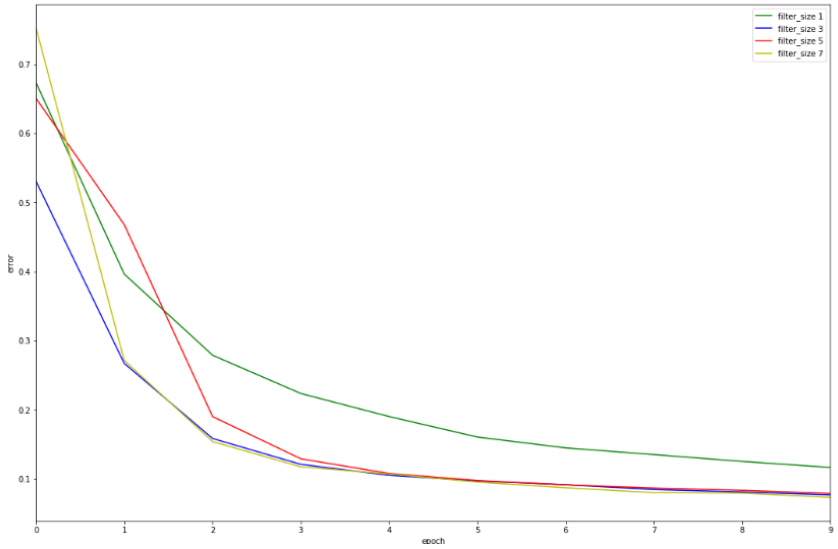

Figure 4: learning curve for different filter size and learning rate 0.001

```
filter_size 1    validation error 0.8432
filter_size 3    validation error 0.7152
filter_size 5    validation error 0.6974
filter_size 7    validation error 0.6276
```

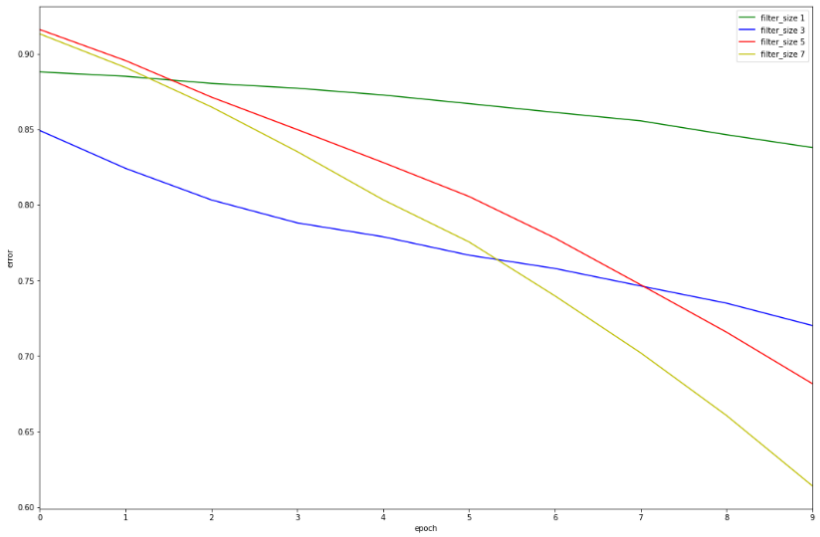

Figure 5: learning curve for different filter size and learning rate 0.0001