Batch Size

Exclaimer:
The whole approach with code can be seen in the notebook. It includes also these results.

Approach

Firstly I checked again what the batch size of a CNN is. It is the number of data-samples that will be put through together in one interation (not epoch) of the network. After each iteration the weights in CNN will be updated. If every data-sample was put through the network, a epoch is finished. So the batch-size on one way determines how 'long' a epoch will take.

The minimum batch-size for every CNN would be 1 and the maximum batch size the number of data-samples available.Minimum and maximum values should be definitly tested.

I selected this six values to represent a good range of results:

- 1
- 8
- 32
- 128
- 256
- 512 (=max of the samples)

Assumptions

I think the higher the batch size, the longer it will take to train the model for each iteration and also the lower the prediction accuracy will be. This could be because if you put all the data-samples in one go through the CNN, the weights will only update once (because there is only one iteration per epoch).

Results

batch size	accuracy	val_accuracy	time per step	time per epoch	epochs needed
1	0.9883	0.9688	~ 7 ms	~ 4 s	22
8	0.9531	0.9531	~ 24 ms	~ 2 s	10
32	0.9766	0.9453	~ 80 ms	~ 1 s	18
128	0.9590	0.9375	~ 285 ms	~ 1 s	30
256	0.9316	0.9531	~ 600 ms	~ 1 s	50
512	0.3965	0.3516	~ 1 s	~ 1 s	5

As you can see the worst accuracy is with bactch size 512 (0.3965 / val: 0.3416). Already after 5 epochs the accuracy couldn't get better, but with ~5 seconds training time it was the fastest.
The best accuracy can be seen with batch size 1 (0.9883 / val: 0.9688), but it took around 88 seconds for the training.

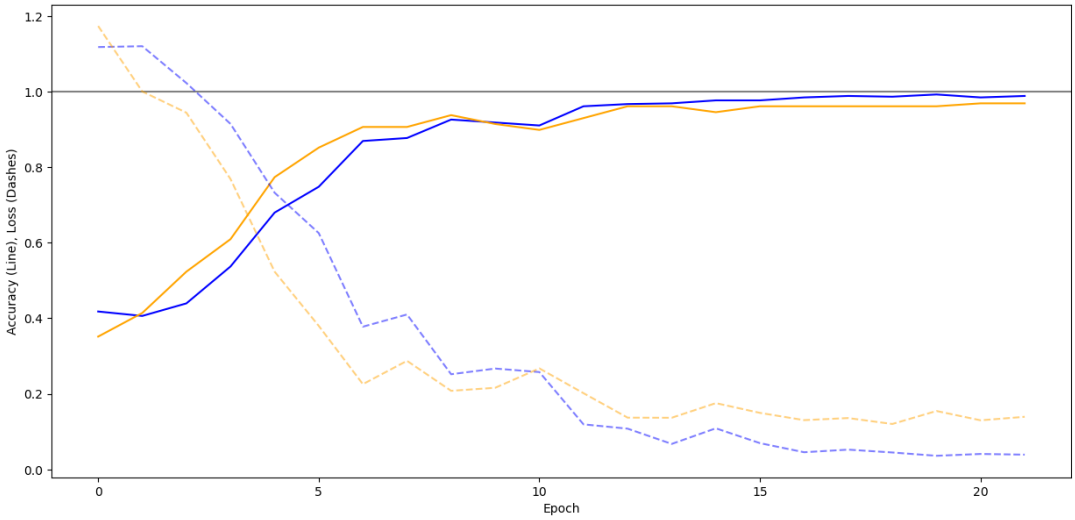
As we can see: the higher the batch size, the lower the accuracy. Except for the batch size = 1, the time to train the model got longer with a higher batch size. But it seems that there is a point, where the model prediciton gets so bad, the model will cancel the training after few epochs, because no better results are shown (batch size 512).

The plots of accuracy and loss of the training process also shows this visually.

The batch size = 1 was very quick for a rather good prediction, in the last ~8 epochs not much progress for better accuracy was made. The plot with the maximum batch size shows that after one epoch accuracy couldn't approve at all.

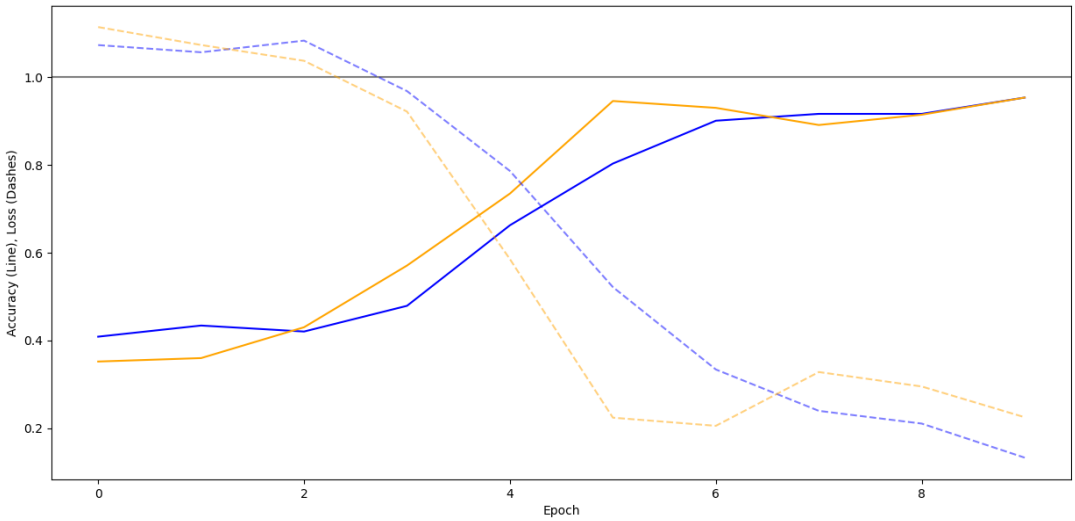
Plots

batch size = 1

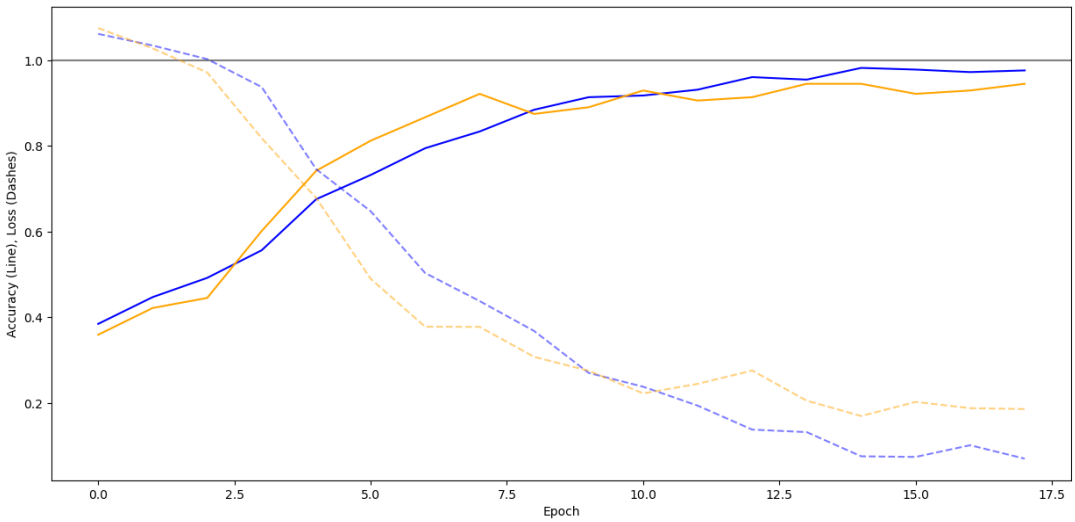


]

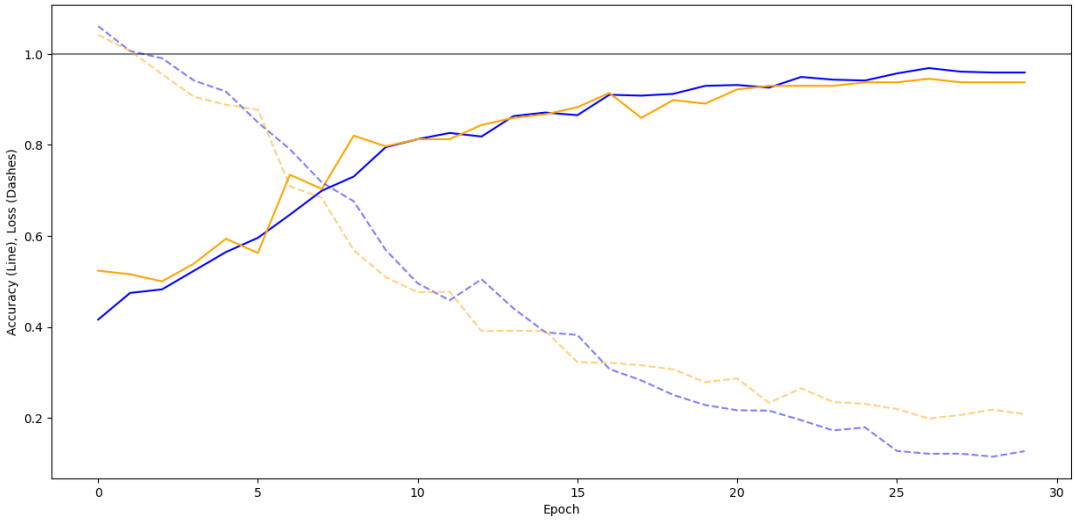
batch size = 8



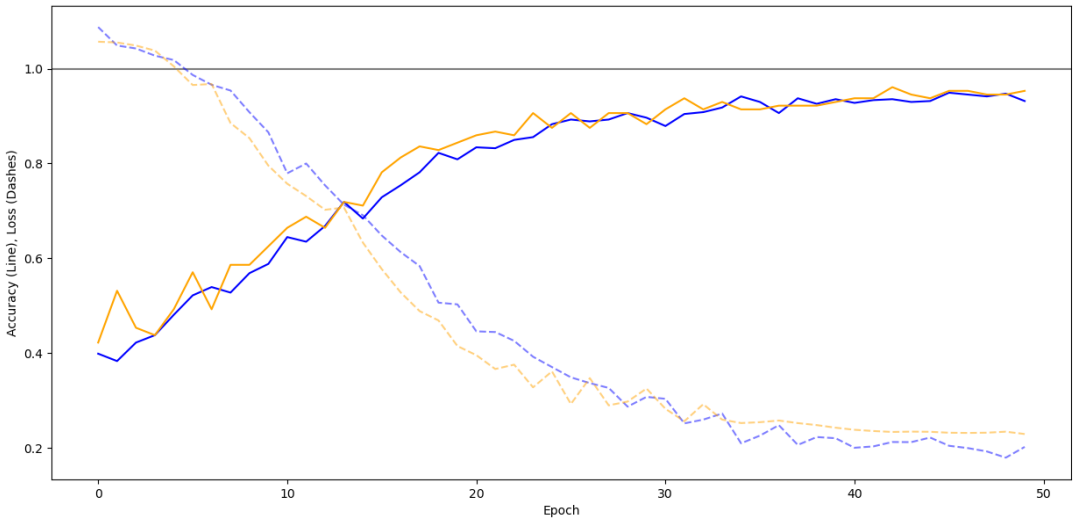
batch size = 32



batch size = 128



batch size = 256



batch size = 512