Practical work 09 Exercise 4 Abtullahu, Rudolf, Gerber

They gave empirical evidence that training with residual connections makes the training of Inception networks significantly faster. It also has a small impact on the performance of the Networks.

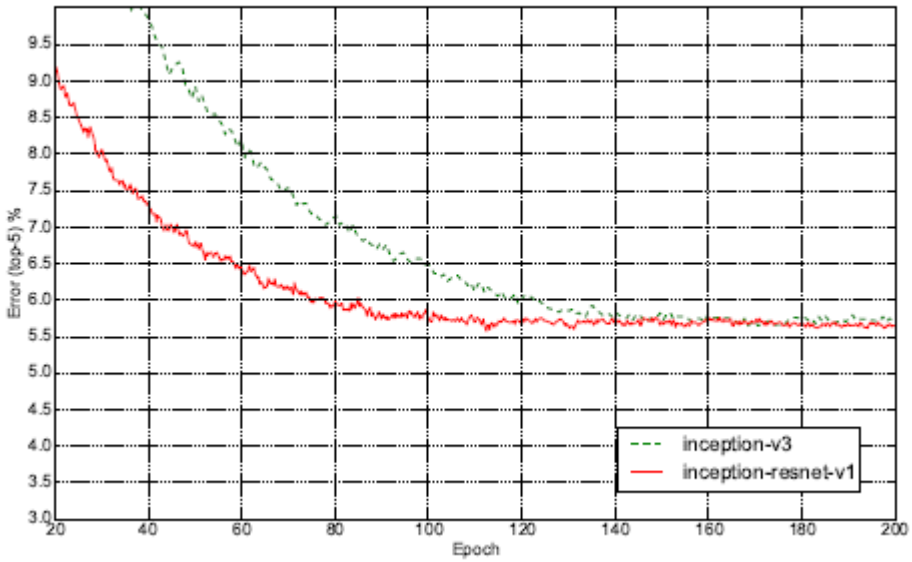

Figure 1 Imporvement of training with residual connections

They also showed that proper activation scaling stabilizes the training of very wide residual Inception networks. It turned out that layers with a large activation size used more GPU.

The Problem is that deeper models are more difficult to optimize. This is clearly not a overfitting problem but rather a optimisation problem. In this case it also makes sence that with this approach the speed of optimisation can be optimized.