

# Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning

## Link

[arXiv](#)

## Summary

- Basically in this paper they incorporated ResNet like residual connection to the inception architecture and compared the performance. ResNet authors claim that residual connections are inherently necessary to train very deep network. But this paper demonstrates that it is not very difficult to train competitive deep network without using residual connection. However adding residual connection greatly improves training time so there is justification to use them.
- When the number of filters exceeded 1000 the residual variants started to exhibit instability and the network just dies early in the training(only zero outputs in the last layer). This could not be prevented by lowering the learning rate or adding batch normalization. They could only prevent it by scaling down the residuals (with 0.1 to 0.3 weight) before adding them to the previous layer activation.
- They achieved similar performance using Inception-ResNet-v2(Inception-ResNet hybrid) and Inception-v4 (pure Inception network).