9.3

In the paper they give clear empirical evidence that training with residual connections accelerates the training of Inception networks significantly. Thus the use of residual connections improves the training speed.

There is also some evidence of residual Inception networks outperforming similarly expensive Inception networks without residual connections by a thin margin

In the paper they also present several new streamlined architectures for both residual and non-residual Inception networks. These variations improve the single-frame recognition performance on the ILSVRC 2012 classification task significantly

If the number of filters exceeded 1000, the residual variants started to exhibit instabilities and the network has just died early in the training, meaning that the last layer before the pooling started to produce only zeros after a few thousands of iterations. Even if the learning rate is very low they still had problems with the instabilities.

They found out that scaling down the residuals before adding them to the previous layer activation seemed to stabalize the training.