SKIP CONNECTION:

The beauty of *deep* neural networks is that they can learn complex functions more efficiently than their shallow counterparts. While training deep neural nets, the performance of the model drops down with the increase in depth of the architecture. This is known as the **degradation problem**. But, what could be the reasons for the saturation inaccuracy with the increase in network depth?

Another possible reason can be vanishing gradient and/or exploding gradient problems. However, the authors of [ResNet](https://arxiv.org/abs/1512.03385) (He et al.) argued that the use of Batch Normalization and proper initialization of weights through normalization ensures that the gradients have healthy norms. But, what went wrong here?

The degradation of training accuracy indicates that *not all systems are similarly easy to optimize.*

One of the primary reasons is due to random initialization of weights with a mean around zero, L1, and L2 regularization.  As a result, the weights in the model would always be around zero and thus the deeper layers can’t learn identity mappings as well.

Here comes the concept of **skip connections**which would enable us to train very deep neural networks.

Skip Connections were introduced to solve different problems in different architectures. In the case of ResNets, skip connections solved the degradation problem that we addressed earlier whereas, in the case of DenseNets, it ensured **feature reusability**.