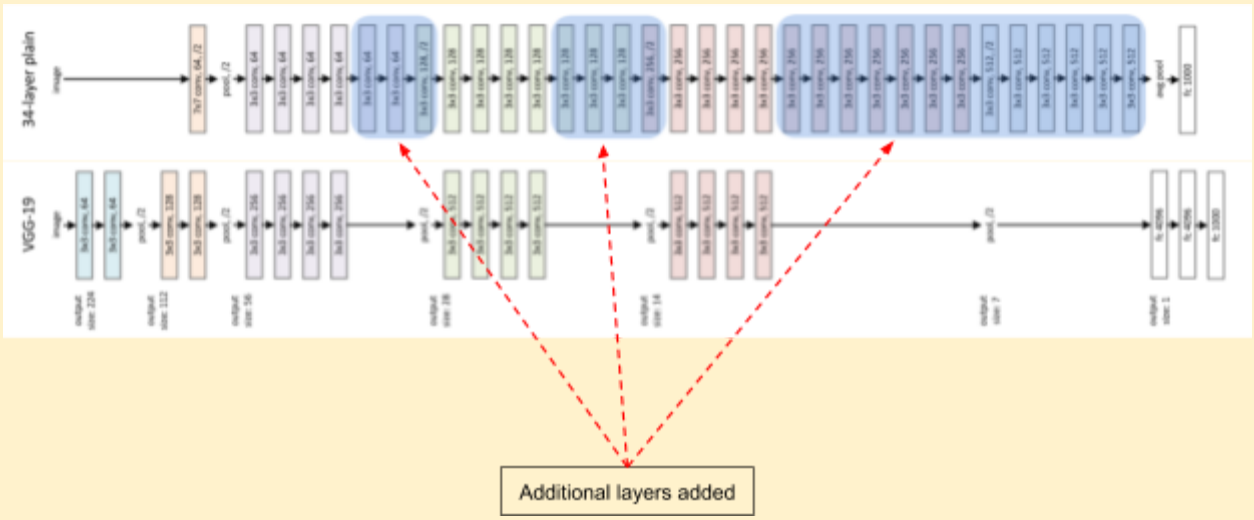
PadhAI: CNN Architectures

One Fourth Labs

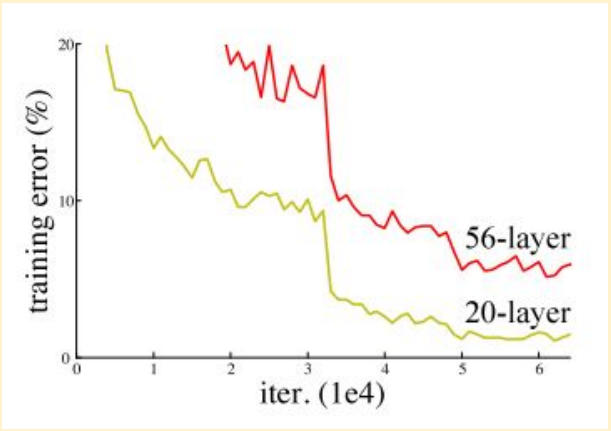
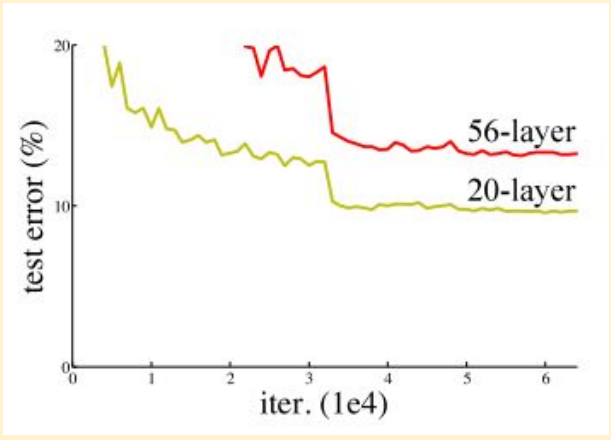
ResNet

What happens if you increase the depth of the network?

1. Let us consider the basic architecture of the VGG-19 network when compared to another 34 layer network architecture

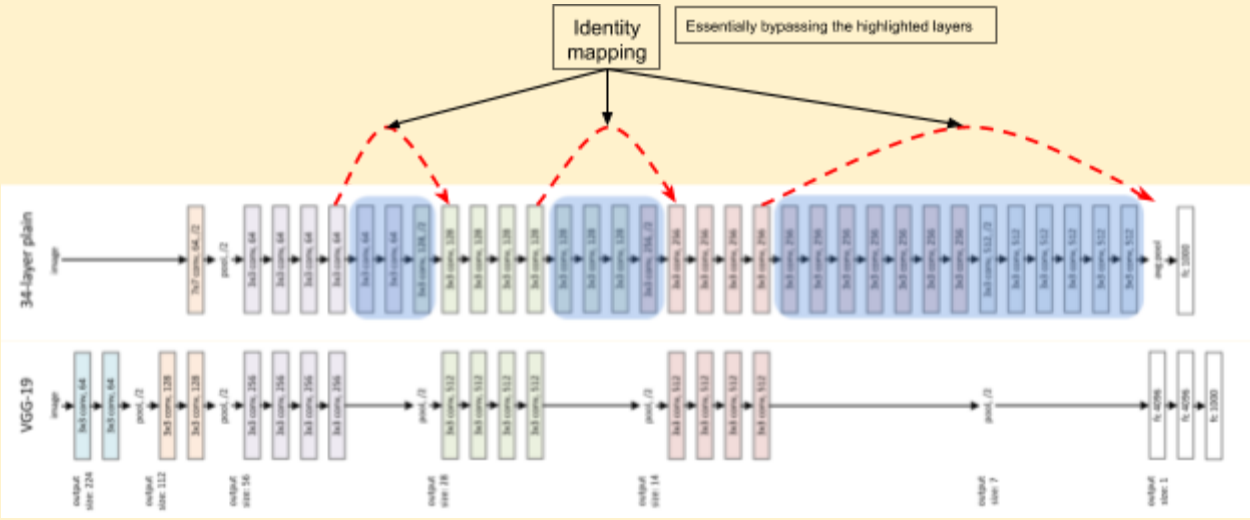


2. The train/test error curves were plotted for some other 20-layer and 56-layer networks

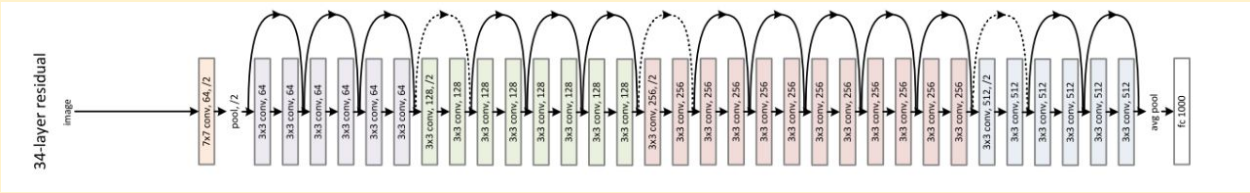
Training curves	Test Curves
	
Unexpectedly, the 56-layer had a higher train error than the 20-layer. It was expected to have overfit the training data, thereby having a lower training error. It was hypothesised that the gradients were not able to flow well through this deeper network.	Here, as predicted, the 56-layer performed worse than the 20-layer due to overfitting.

3. Now, when comparing the 19 and 34-layer networks, we see that at the very least, the 34-layer network should be able to match the performance of the 19-layer network.

4. Matching the performance could be done by bypassing the additional layers in the 34-layer network by using Identity Mapping.
5. Identity mapping refers to learning filter values such that the output is preserved identical across a finite number of layers, till it reaches the target layer. Basically cloning the layer output till required.



6. However, it wasn't able to match the 19-layer's error. This implies that the information from the input is getting highly morphed and by the time we reach the output, it is highly transformed.
7. A simple solution would be to keep passing the input information repeatedly in stages.
8. To attempt this, they tried the Residual Network or the ResNet



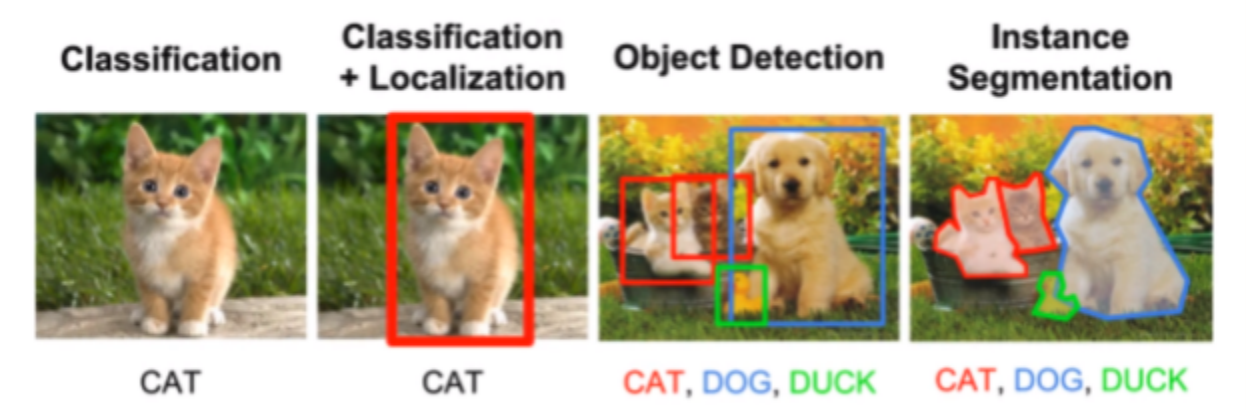
9. In the ResNet, **every two layers, we pass the input given to the first layer along with the output obtained at the second layer.**
   a. Input: x_1
   b. Output: $x_2 = f(x_1) + x_1$
   c. Output after two layers: $x_3 = f(x_2) + x_2$
10. This helped the gradients to flow back better and the training to improve
11. It is called a Residual Network because at every stage, there is a residue of the input which is passed once again with the output.
12. Using this technology, they were able to train very deep Neural networks of up to 151 layers.

PadhAI: CNN Architectures

One Fourth Labs

---

13. The ResNet showed remarkable performance among the various tasks



14. It was the winner among the 4 main tasks across the following datasets
    a. ImageNet Classification **(ResNet-151)**
    b. ImageNet Localization **(ResNet-101)**
    c. ImageNet Detection **(ResNet-101)**
    d. Coco Detection **(ResNet-101)**
    e. Coco Segmentation **(ResNet-101)**
15. Some of the popular ResNets are **(ResNet-51, ResNet-101 and ResNet-151)**