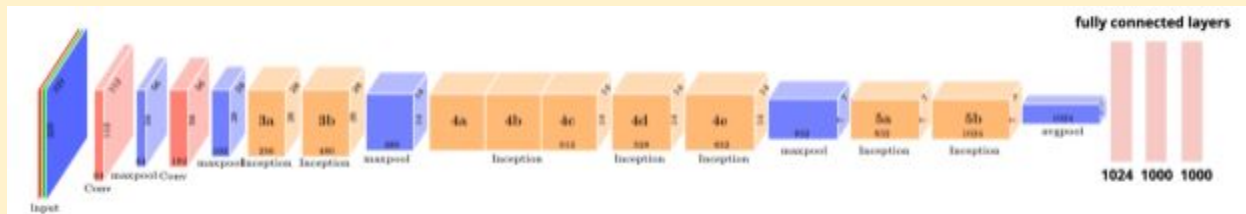


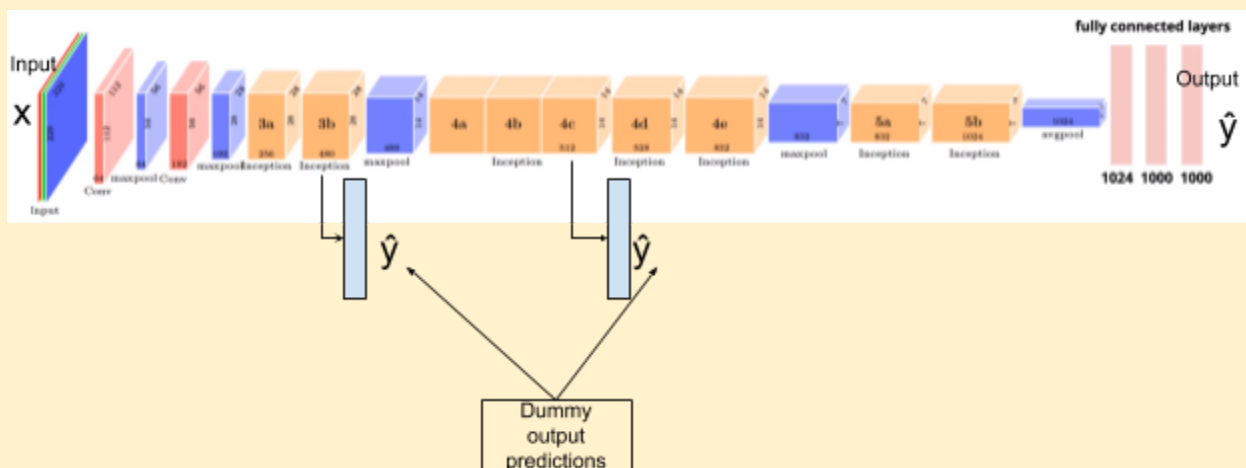
### Auxiliary Loss for training a deep network

Can auxiliary loss help to train the network better?

1. Let's look at how GoogLeNet responds to the 4 problem points from the previous CNN architectures



2. **Increase choice of filters:** Parallel convolutions/max-pooling
3. **Reduce number of parameters:** Average Pooling
4. **Reduce number of computations:** 1x1 convolutions
5. **Make a deeper network:** Has 22 layers as opposed to VGG19's 19 layers.
6. Now, since it is a very deep network, there is a possibility for vanishing gradients to occur when backpropagating the Loss. This is mitigated using a technique called Auxiliary Loss



7. In addition to the final output prediction, we are also trying to make partial predictions from the above specified regions in the network.
8. We compute the loss at the final prediction and at both the dummy predictions.
9. Now we can backpropagate from the final loss or from the dummy-losses obtained, thereby shortening the effective depth of the network and lowering the chance of vanishing gradients occurring.
10. Some interesting points to note about GoogLeNet
  - a. 12x less parameters than AlexNet
  - b. 2x more computations than AlexNet
  - c. Improved performance on ImageNet