some notes..

(Convolution & Pooling)

- Conv layer can be viewed as a special case of a fully connected layer when all the weights outside the local recepti ve field of each output neuron equal 0 and kernel parameters are shared between neurons.
- Conv layer works the same way for every input patch.

- Pooling layer provides traslation invariance.
- Pooling layer can reduce the spatial dim of input volume, when used with stride > 1.

(Backprop - Conv)

- Backprop for conv layer first calc the grad as if the kernel parameters were NOT shared and then takes a sum of gr ad for each shared parameter.

(# parameters)

- Assuming 10*10*3 color image input to a 2 conv layers with kernel size 3*3 with 10 and 20 filters applied respecti vely.

for the first layer: (3*3*3 + 1) * 10
for the second layer: (3*3*3 + 1) * 20

A total of 2100 parameters.

(ResNet)

- Even with sufficiently deep NN, identity fn can NOT be easily learnt; due to problems like dying grad or curse of d imensionality.

- In practice, by incerasing # layers the accuracy will start to saturate & eventually degrade. This is NOT overfitting! This is the Degradation problem.

(Residual Connections)

- It's easier to optimize the residual mapping than to optimize the original input mapping.
Even in the extreme case, where the identity (i.e input) mapping is optimal, it is still easier to push the residual to zer o than to fit identity mapping by a stack of non-linear layers.
- Skip Connections allow the network to more easily learn identity-like mapping.