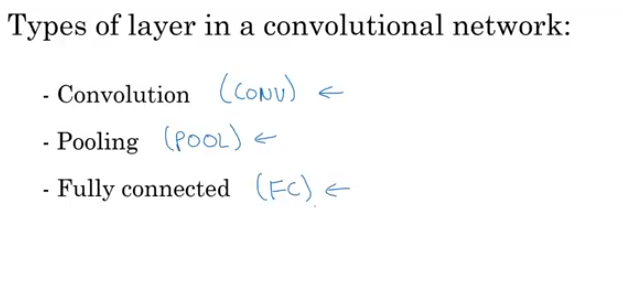
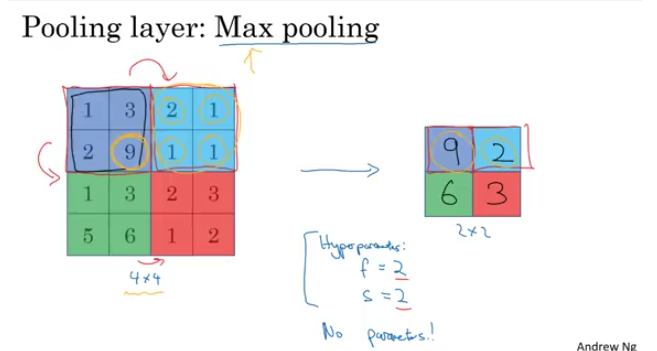


General Trend**: Image dimension decreases as go deeper in to NN towards final prediction**

**While number of channels increases as go deeper in to NN towards final prediction**

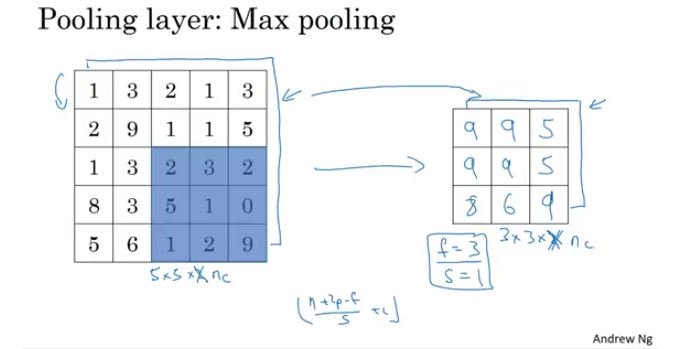
****

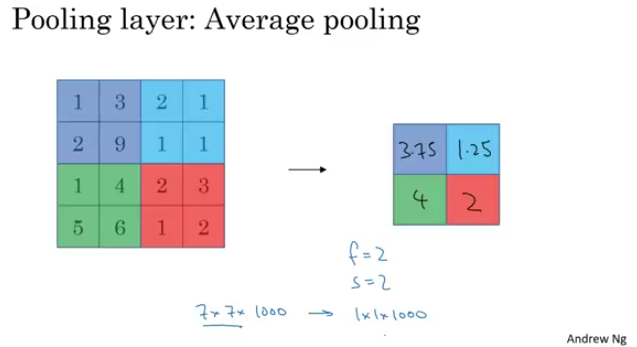
**Pooling layers added in terms of detection of features robustness and speed.**

****

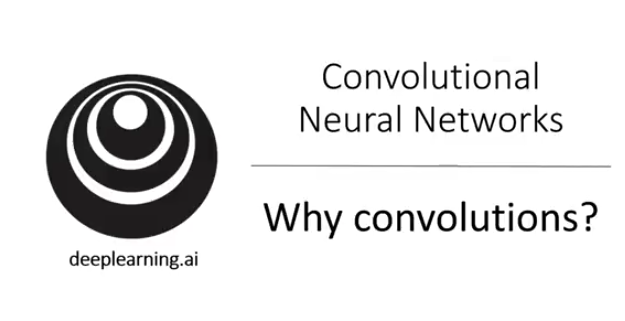
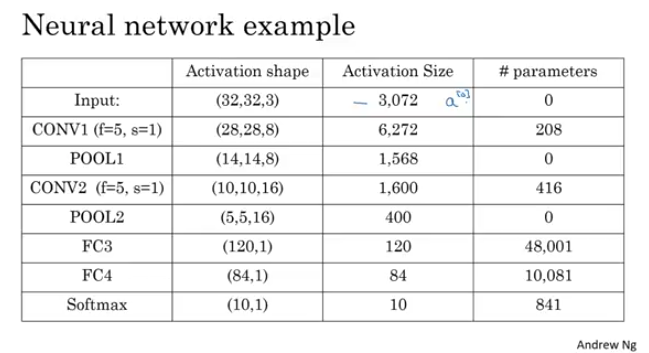
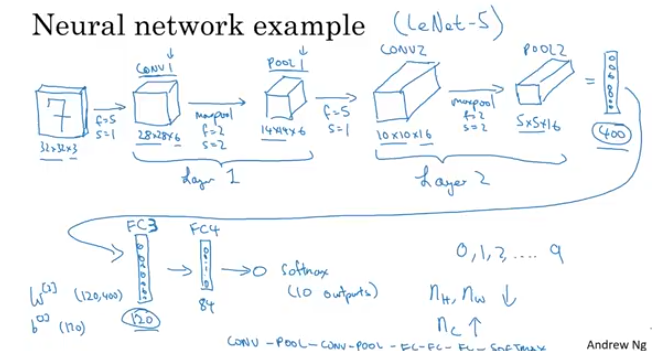
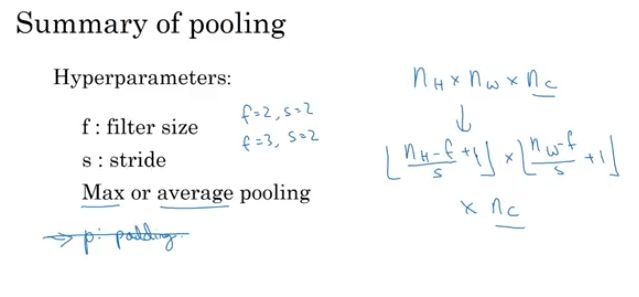
**Independent layer in output for each layer in input**

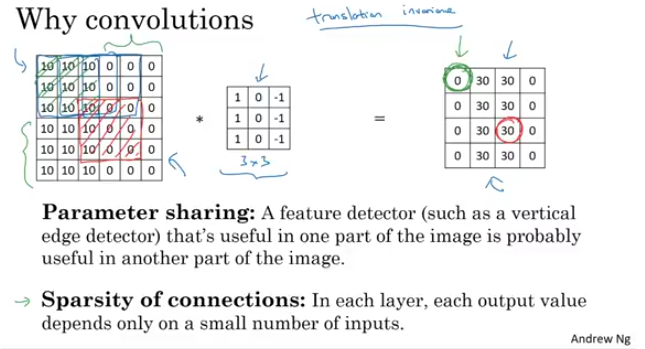
Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network, and hence to also control overfitting. The Pooling Layer operates independently on every depth slice of the input and resizes it spatially, using the MAX operation

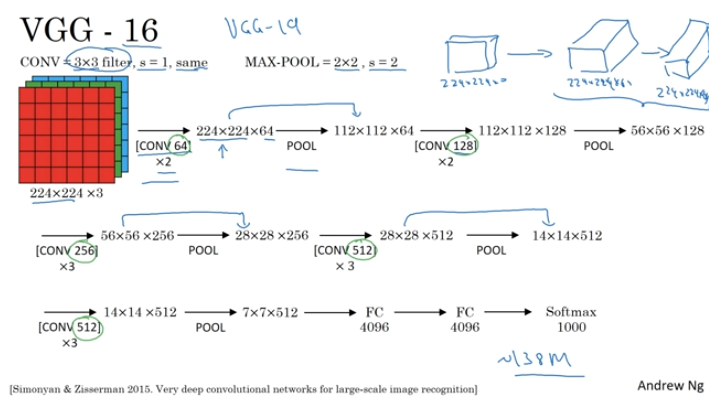
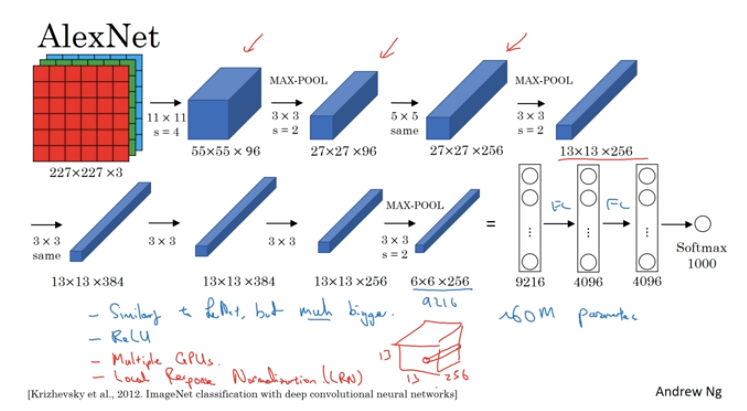
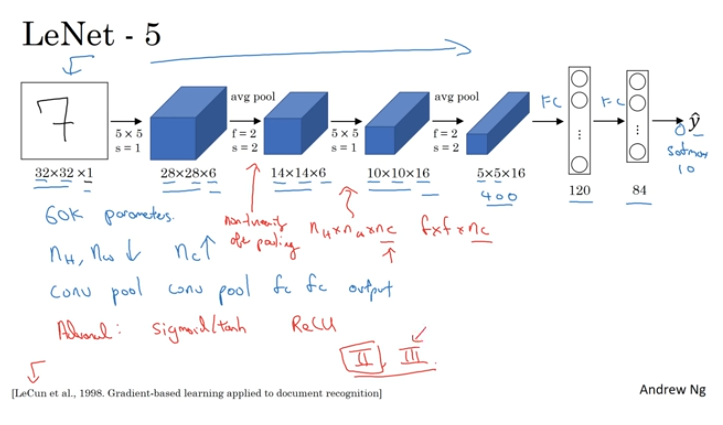
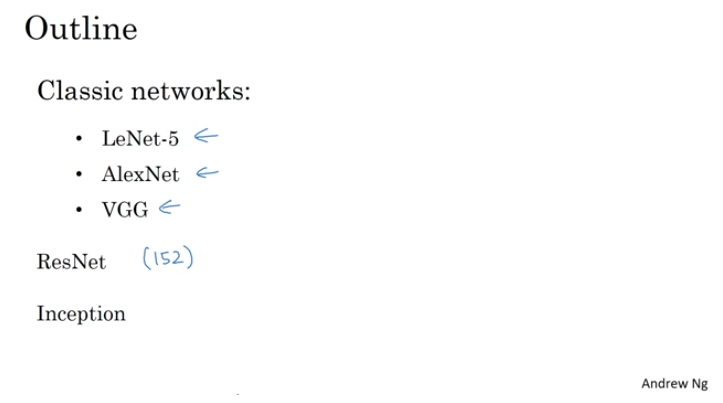
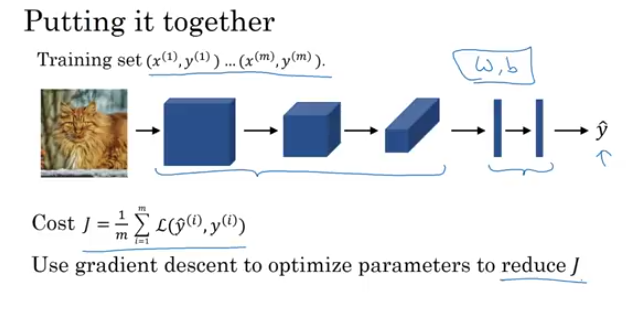
****

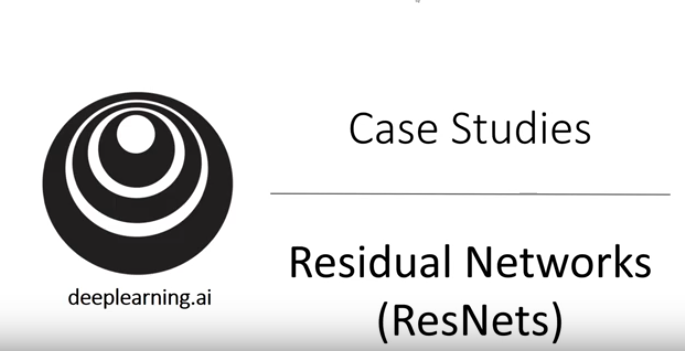
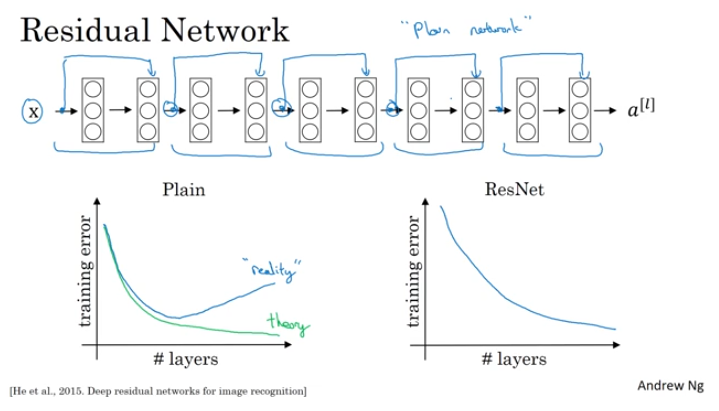
****

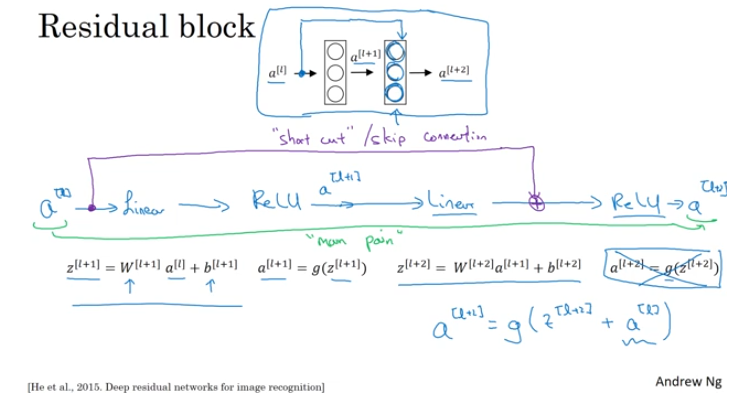
**Max pooling is used often rather to average poling.**

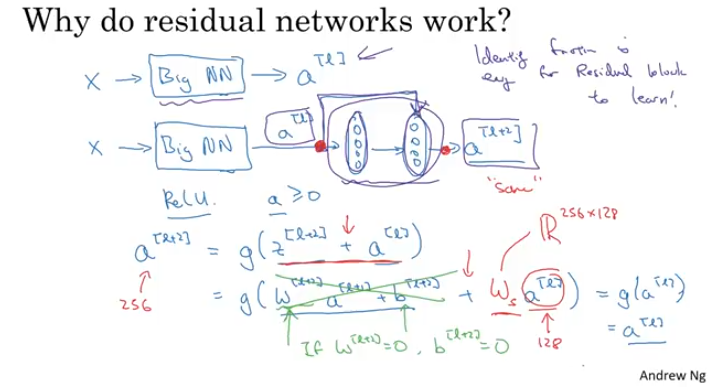
**No parameter to learn during Backprpagation.**

****

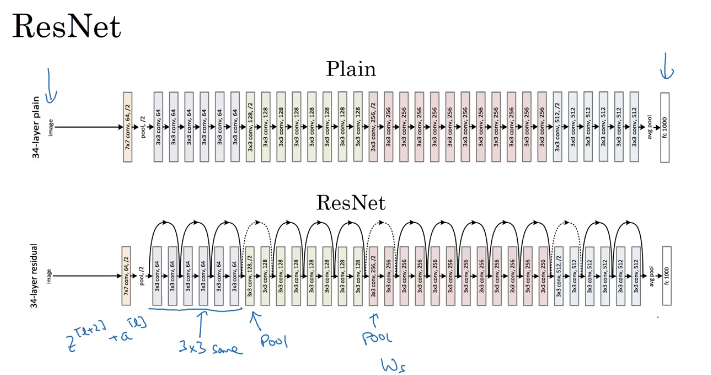
****

**Work very much better in case very deep network and remove the problem of vanishing or exploding gradient decent **

****

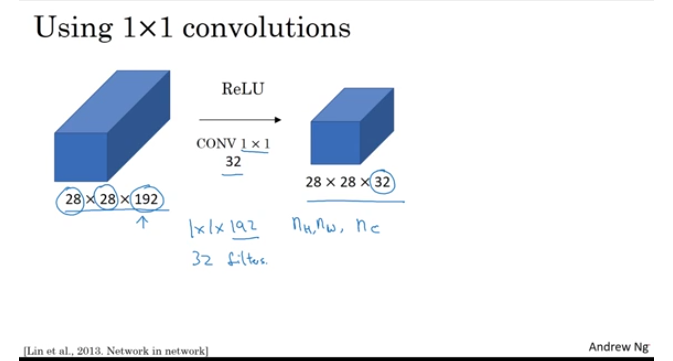
****

**Parameter shrinks due to regularization so this complete residual unit work as a identity unit.**

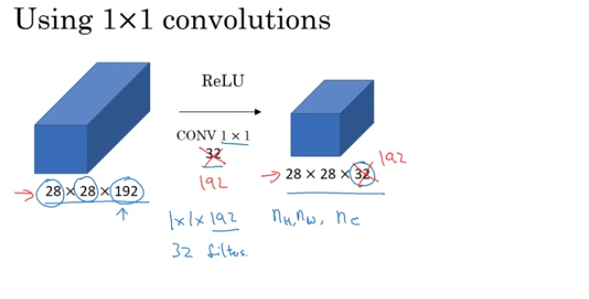
**To make addition easy for a(l) and a(l+2) ,same convolutions are used.**

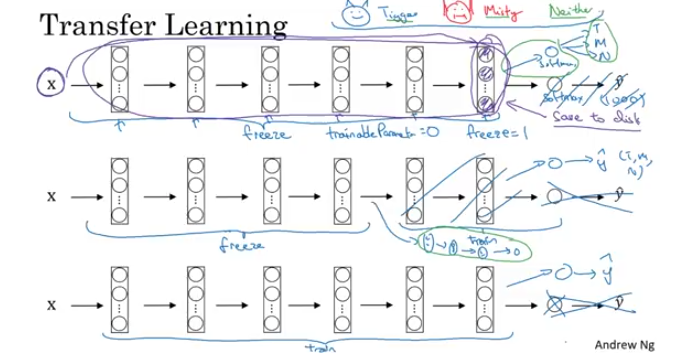
****

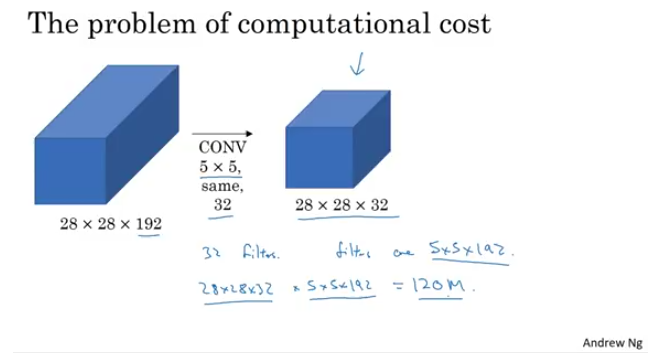
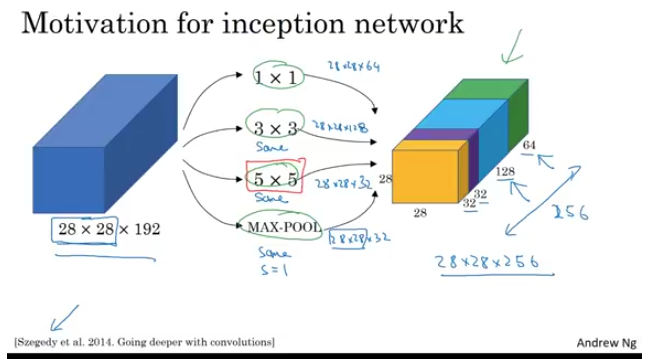
**Used to shrink number of channels**

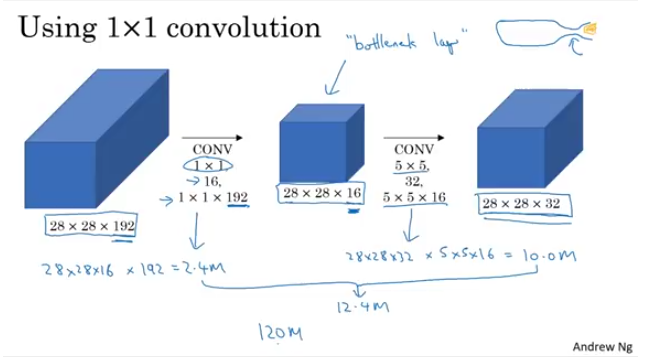
****

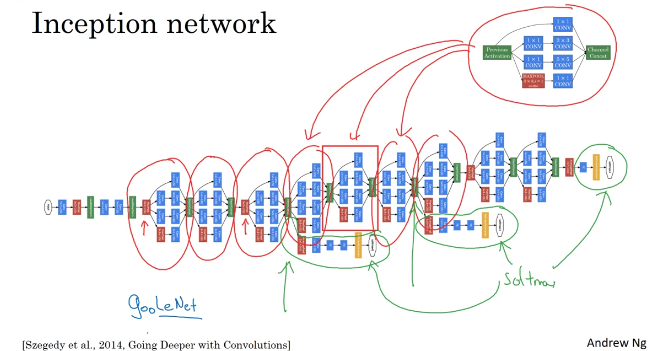
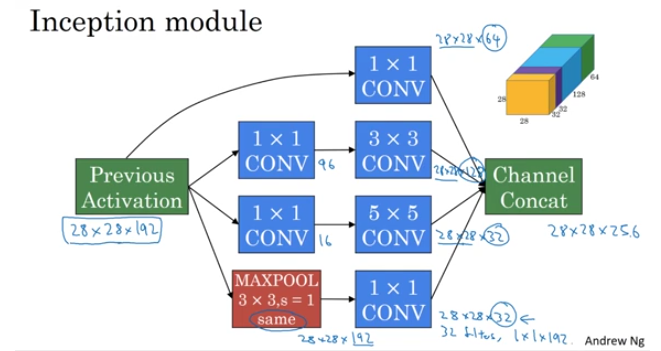
**Just to add only non linearity**

****

****

****

****

****