Filters In CNN are same as weights in ANN, they are learnt by forward and backward propagation.

But in the pooling layer, that filter is fixed and doesn't learn anything. It saves computation and detect only major features. Padding is not or rarely used in pooling because it zero and nothing will be gotten in max pooling because it try to get only max value.

There are two conventions, some consider conv layers is one and pooling as other layer, and some consider conv and pooling as single layer.

CNN are faster, every output only depends on few inputs like in filter of 3x3 one output location is depends only on 9 values but in fully connected layer, one output depend on every input.

LeNet, AlexNet, VGG

Very deep networks vanishing and exploding gradients, so in ResNet is better in deep networks. In this after every two layers, previous layers results are added in second next layer just before applying activation.

Network in Network is 1X1 convolution. Used as bottle neck layer in Inception to reduce computational cost.

In Inception module, we take multiple types of filters in a single layer and stack them up in single volume. In Inception network there are many inception modules.

In MobilNet Network. Depth wise separable convolution include Depth following pointwise convolution.

MobileNet v1 there is depthwise and pointwise convolution and in MobileNet v2 in addition to, there are residual connection and expansion layer. Expansion and pointwise convolutions are similar, difference is that, in expansion it increases the volume by using greater number of filters and in pointwise convolutions it reduces by using lesser filter quantity, it also called projection.

They are very less in computational cost.

EfficientNet.

ImageNet is a dataset that has almost 14 million images and 1000 classes.

Ensembling, training several different networks and averaging their y hat or output.