* How to predict the different?
* CNN:
* How it works? How algorithm knows?
  + Is it a shape? Is it a region?
  + Have many filters and these filters allow to extract the features from different part of the image.
  + We use filters and these filters look for the specific features in the image,
* Convolution: process of combining two functions to produce third.
  + Image \* kernel = feature map (overalay filter to the image and multipy and add. Then shift one step to the right.
  + 5\* 5 \* 3\*3 = 3\*3
  + Feature map size: n-f+1 = m (n= image, f = kernel, m = feature map)
  + Feature Detectors: learn how what the image is bit by bit.
  + Convolutions on color images:
  + (n\*n\*nc) \* (f\*f\*fc) = (n-f+1) \* (n-f+1)\* nf
  + No. of filters = no. of features
  + Kernel size and depth (parameter) control size of cov filter:
  + Kernel size: has to be odd numbered and squared.
  + Why can’t have even numbered: No center point, mess up the symmetry, distortion to the lower . Do not work
  + Depth: no. of kernels
* Padding: manipulate the feature map size, put zero around the matrix. Why to use: for the deep networks for complex images, preserve image size, touches more so can extract information from the edges
* Stride: how many steps we take while convolving. Larger stride: less overlap, use to control feature map,
* (n+2p-f)/s +1
* Activation layer (Relu): change negative values to 0 and leave all positive values alone. Same in convolution layer.
* Why activation function;

Introduce non-linearity:

* + Max pooling: reduce the size of feature map. Dimension reduction.reduce no. of parameters, faster
  + Also called subsampling or downsampling.
  + Use filter 2\*2 and stride of 2 with no padding.
  + Closer pixels are highly correlated.
  + Fully connected layer: flatten to single vector. Compiles the data to form output. Way to learn neural network
  + Final layer: SoftMax layer: probabilities.
  + 