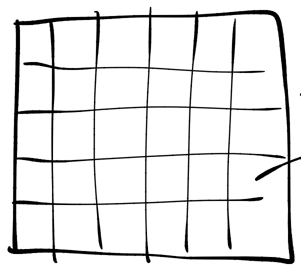
LECTURE 13

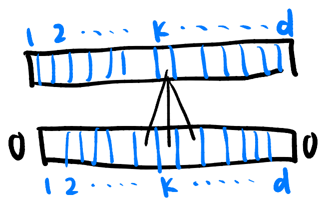
convolutional neural network
(CNN, conv. net)



- each square is a pixel
- pixel is local intensity

2D image (100 x 100)

start from 1D



output y

input x

2 adjacent layers of network

both x & y are 1D maps
k: index position of map

$$y_k = w_{-1} x_{k-1} + w_0 x_k + w_1 x_{k+1}$$

for all $k = 1, 2, \ldots, d$

kernel, filter → 3 linear parameters

w_{-1}	w_0	w_1

to be learned

shared over all $k = 1, \ldots, d$

sliding window, translation invariant
zero padding at boundary

patterns can appear in different locations

to be followed by adding bias, rectification

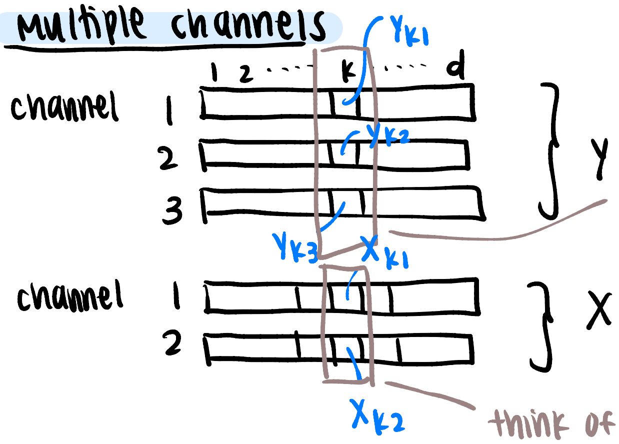
bigger filter

w_{-3}	w_{-2}	w_{-1}	w_0	w_1	w_2	w_3

linear transformation can happen btwn any adjacent layers

$$y_k = w_{-3} x_{k-3} + w_{-2} x_{k-2} + w_{-1} x_{k-1} + w_0 x_k + w_1 x_{k+1} + w_2 x_{k+2} + w_3 x_{k+3}$$

Multiple channels

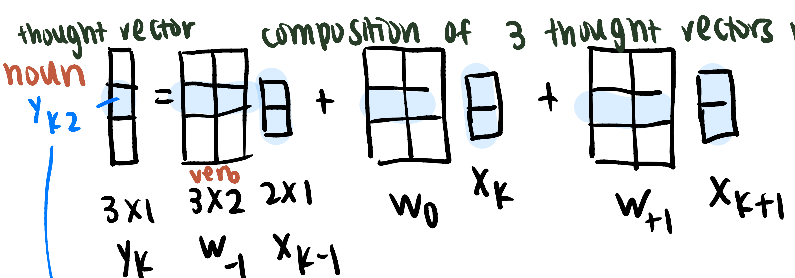


think of y_k as a 3D vector

think of x_k as a 2D vector

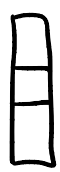
$y_k = W_{-1} x_{k-1} + W_0 x_k + W_1 x_{k+1}$

W is a 3x2 matrix



computed by combining all the channels below

+ bias

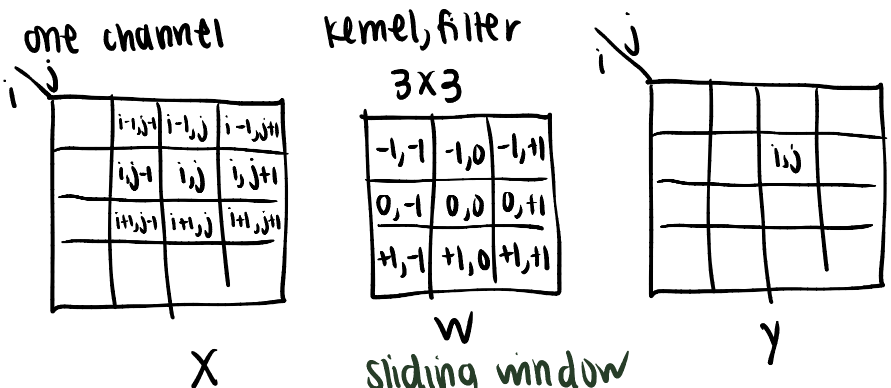


3x1 vector

+ Rectification (applied elementwise)

Back to 2D

need to generalize formula & sliding window to 2D:

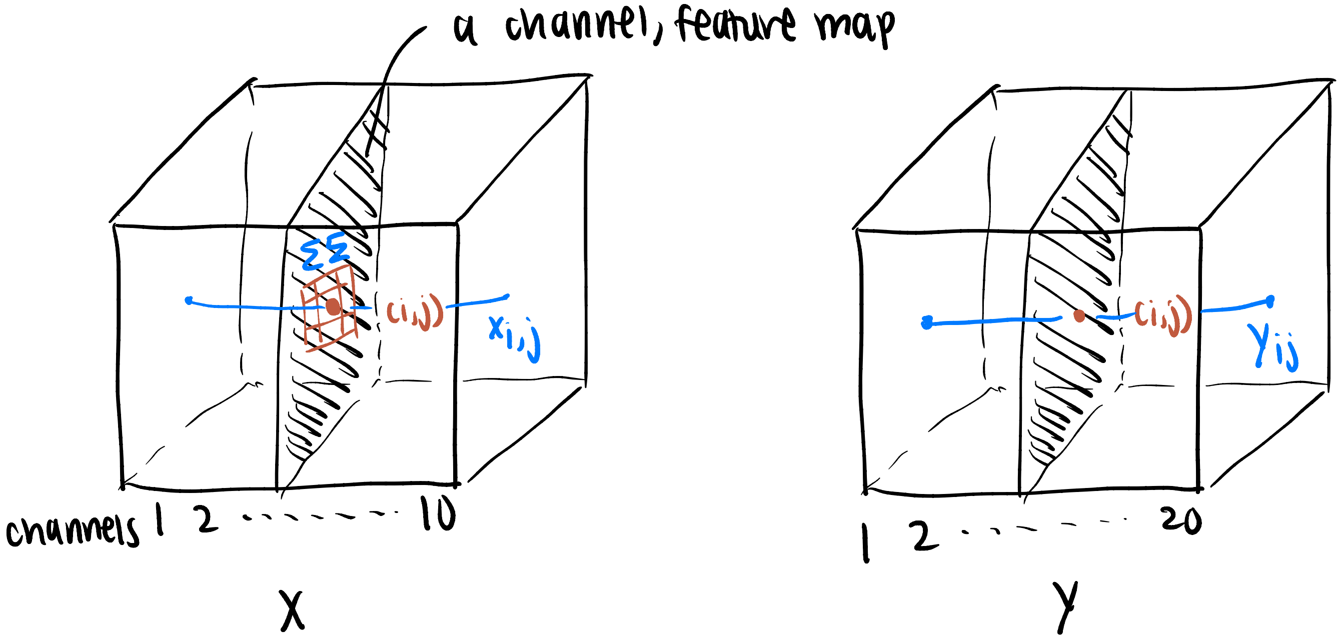


$$y_{ij} = W_{-1,-1} x_{i-1,j-1} + W_{-1,0} x_{i-1,j} + W_{-1,+1} x_{i-1,j+1} + W_{0,-1} x_{i,j-1} + W_{0,0} x_{i,j}$$
$$+ W_{0,+1} x_{i,j+1} + W_{+1,-1} x_{i+1,j-1} + W_{+1,0} x_{i+1,j} + W_{+1,+1} x_{i+1,j+1}$$

+ bias, rectification

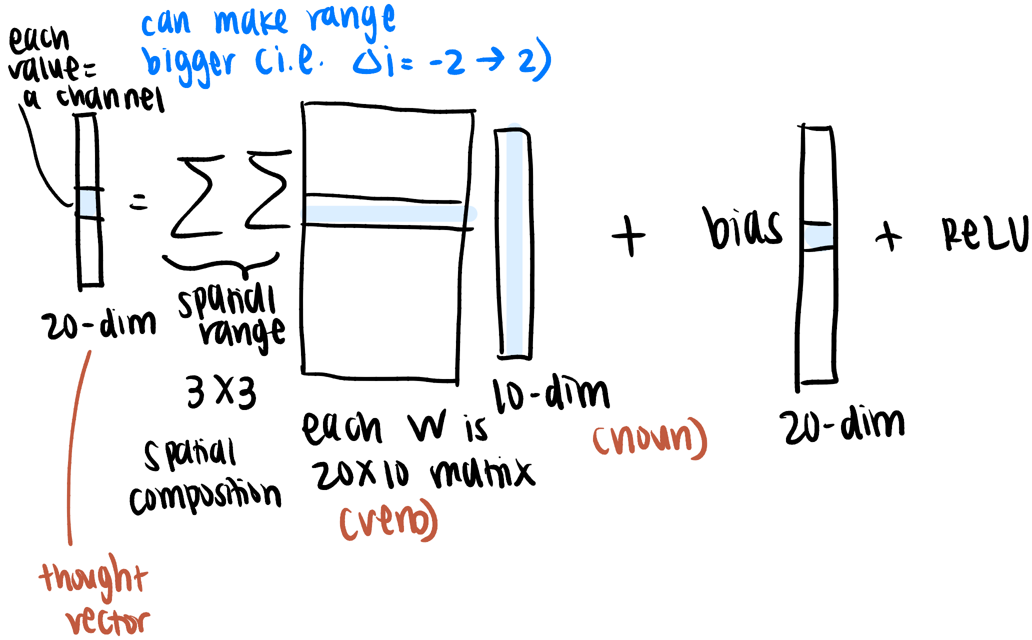
can make this bigger (i.e. 5x5, 7x7...)

now generalize multiple channels to 2D



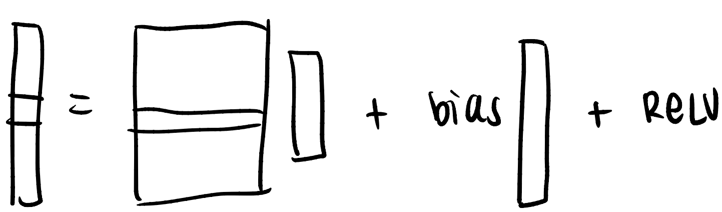
$$y_{ij} = \sum_{\Delta i=-1}^{1} \sum_{\Delta j=-1}^{1} w_{\Delta i, \Delta j} \, x_{i+\Delta i,\, j+\Delta j}$$

can make range bigger (i.e. $\Delta i = -2 \rightarrow 2$)



1 x 1 conv., no spatial pooling

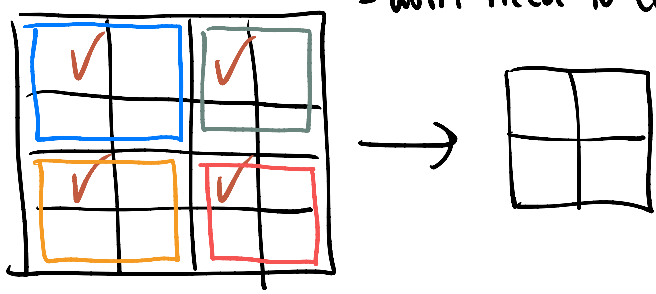
$$\underset{\text{20-dim}}{y_{ij}} = \underset{20\times10}{w_{0,0}}\, \underset{\text{10-dim}}{x_{ij}}$$



network within network

Subsampling

↳ makes map smaller, computation faster
  - don't need to calculate redundant observations



4x4 reduced to 2x2

stride = 2

OR max pooling (find maximum in each block)

OR 2x2 filter (may be better than direct subsampling)

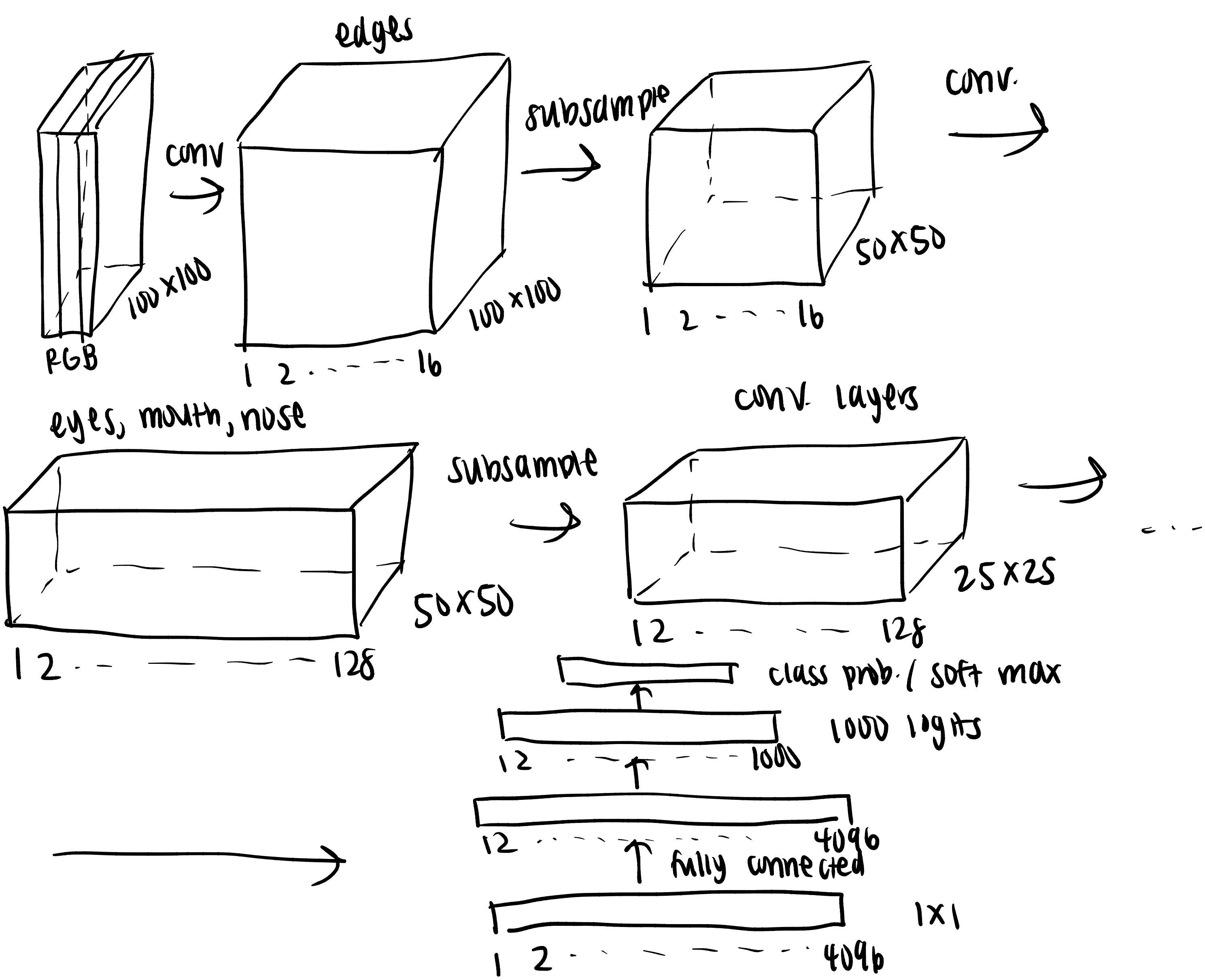
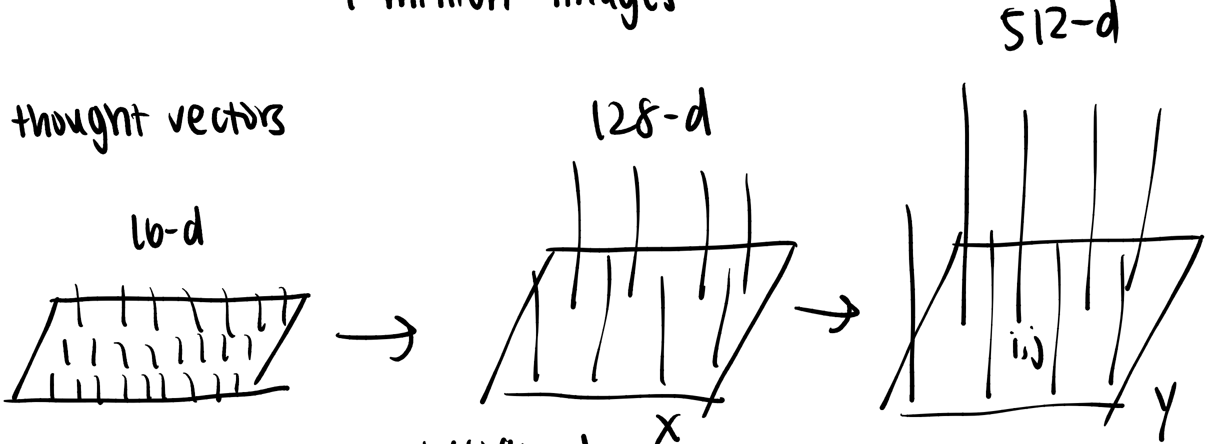


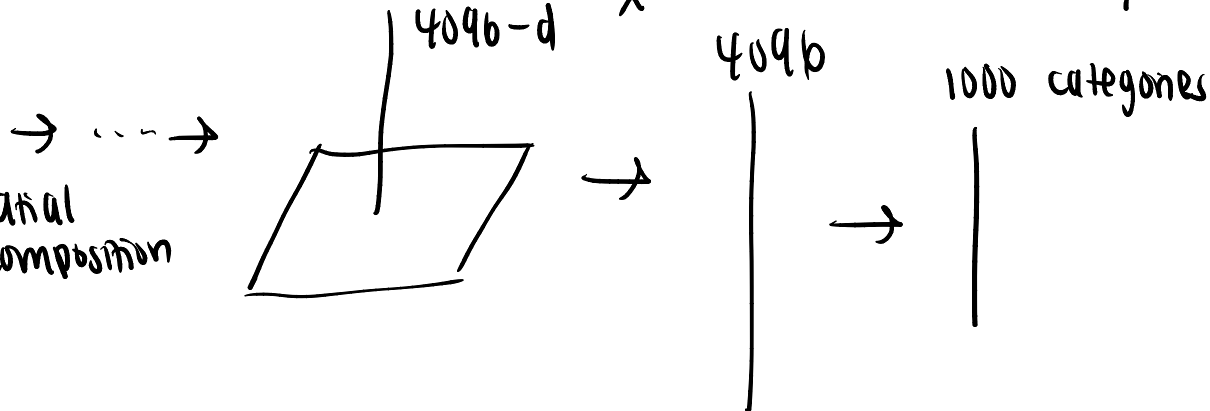
image net

1000 categories
1 million images

thought vectors



spatial composition



$$y_{ij} = \sum_{\Delta i} \sum_{\Delta j} w_{\Delta i \Delta j} \, x_{i+\Delta i,\, j+\Delta i}$$