Convolutional Neural Networks

Convolutional Layer

1. Fully connected layer
    a. 32 x 32 x 3 image should be stretch to 3072 x 1 vector
        i. input: 3072 x 1
        ii. weights: 3072 x 10
        iii. activation: 10 x 1
    b. Cannot preserve the spatial structure
2. Convolutional layer
    a. Convolve the filter with the image
        i. slide over the image spatially, computing dot products
        ii. filters always extend the full depth (RGB = 3) of the input volume
    b. One number is resulted from taking a dot product between the filter and a small chunk of the image
        i. chunk has a same size with filter
        ii. e.g., 5 x 5 x 3 filter → 75-d vector dot product + bias
    c. Filter slides over all spatial locations of the image
    d. Activation map is the collection of the output numbers
        i. e.g., 32 x 32 x 3 image with 5 x 5 x 3 filter generates 28 x 28 x 1 activation map
        ii. the depth of the activation map is equal to the number of filters
    e. ConvNet is a sequence of the convolutional layers interspersed with activation function
    f. Stride
        i. the number of pixels to jump while sliding
        ii. not every number is applicable for stride
            - e.g., 7 x 7 input image, 3 x 3 filter 일 때, stride 3은 사용불가

iii. output size = ((N - F) / stride) + 1

- N: input size
- F: filter size

iv. intuition for using stride

- down sampling을 위함
- 큰 값의 stride를 사용할수록 더 많이 down sampling 됨
- parameter 의 수에 영향을 줌

g. Zero padding

i. zero pad the border

ii. output size = ((N + 2 x pad - F) / stride) + 1

iii. maintain the input size

- e.g., 7 x 7 image, 3 x 3 filter, stride 1, pad 1 이면 output size = ((7 + 2 x 1 - 3) / 1) + 1 = 7
- e.g., 32 x 32 x 3 image, 5 x 5 filters 10, stride 1, pad 2 이면 output size = ((32 + 2 x 2 - 5) / 1) + 1 = 32
  - number of parameters = (5 x 5 x 3 + 1) x 10 = 760
  - 필터 한개의 크기는 5 x 5 x 3 이고 + 1은 bias term 이므로 10개의 필터에 대한 파라미터 수는 760개
- common to use conv layers with stride 1, filters of size F and zero padding with (F - 1) / 2 to preserve the size spatially

h. Shrinking problem

i. input convolved repeatedly with filters shrinks volumes spatially

ii. shrinking too fast is not good b/c lose out some information

3. Pooling layer

a. Make the representations smaller and more manageable

i. use fewer parameters by down sampling

b. Operate over each activation map independently

c. Depth is not changed

d. Max pooling

i. can use different filter size and strides

ii. most commonly used

- 해당 지역에서 가장 튀는 값을 잡아내는 것이 그 지역의 특성을 잘 반영할 수 있음
- 평균을 내는 방식은 비슷한 경향을 갖게 될 수 있음

e. Stride can be used instead of pooling

i. 최근 연구에서는 fractional stride 등 다양한 stride를 이용하여 pooling 대신 down sampling에 사용하기도 함

4. Fully connected layer

a. Contain neurons that connect to the entire input volume as in ordinary neural networks

b. Take the output of convolutional neural network

i. stretch out to 1-d vector