컴퓨터비전특론 4th week summary

[Lec06: Convolutional Neural Networks

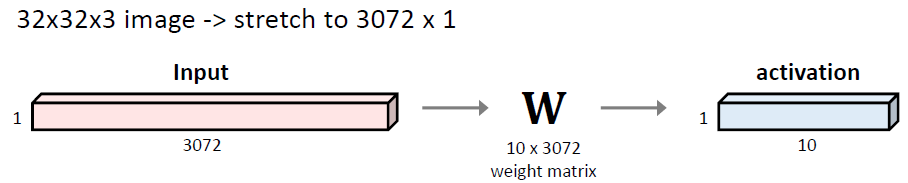
ConvNets가 사용되는 분야

: Classification, Retrieval(영상을 입력하면 비슷한 종류의 영상을 출력), Detection(Bounding box 형성 & box 안의 물체를 labeling), Segmentation(픽셀 단위의 labeling), Action recognition, Human pose estimation, Deep reinforcement learning, Image captioning, Style transfer

ConvNet

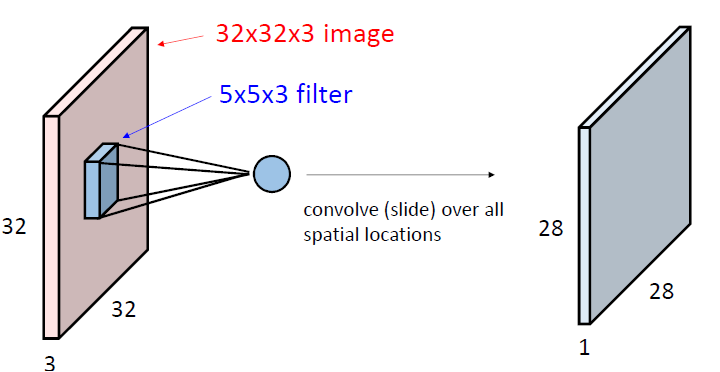
Defferentiable function을 이용하여 data를 다른 형태로 변환하는 sequence of layers

- Fully-connected layer



Fully connected layer는 Image를 1D vector로 만들어 vector 안의 모든 원소가 output vector의 원소 하나를 계산하는 데에 이용된다. 그러나 이 경우 이미지 데이터에서 특정 픽셀과 매우 유사한 특성을 가질 가능성이 높은 주변 픽셀들과 멀어질 수 있는 문제가 있으며, 파라미터 개수가 과도하게 많아지는 문제가 있다.

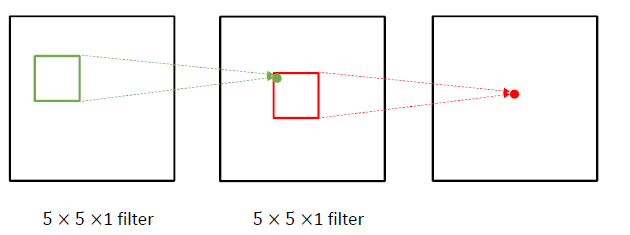
- Convolutional layer



Convolutional layer는 2D image를 입력값으로 받아 이미지 데이터의 spatial structure를 반영한 output을 얻을 수 있다. Filter 하나는 input image를 왼쪽 위부터 시작해서 filter 위치를 이동시키면서 filter와 중첩된 image 부분의 값에 대한 weighted sum을 계산한다. 한번의 계산마다 하나의 값이 계산되며, 이에 대한 수식은 이다. 예를 들어 32 x 32 x 3 image에 5 x 5 x 3 filter를 적용하면 28 x 28 x 1 크기의 output을 얻을 수 있다. 하나의 필터가 가지고 있는 숫자들이 weight를 의미하므로, 하나의 output feature map이 만들어질 때 모든 픽셀이 동일한 weight를 share한다. 위의 예시에서 6개의 filter를 사용하면 output으로 28 x 28 x 6 size의 feature map을 얻을 수 있다.

- Receptive field

Output feature map의 하나의 픽셀을 만들기 위해 input image의 얼마나 많은 영역이 cover되어야 하는지를 뜻한다.



예를 들어 위의 예시에서 가장 오른쪽 이미지의 픽셀 하나(빨간 점)를 만들기 위해 첫번째 이미지의 5 x 5, 두번째 이미지의 5 x 5가 사용되었다. 이 경우 빨간 점의 effective receptive field는 9 x 9 (=5+5-1)이다.

- Stride

Convolution 과정에서 filter가 이미지 전체를 이동할 때, 한번 이동할 때 움직이는 픽셀의 칸 수를 뜻한다.

- Output size

(N-F) / stride + 1, N: input size, F: filter size.

Output size는 정수가 되어야 하므로 이를 고려하여 stride를 결정해야 한다.

- Padding

Image 가장자리의 정보까지 모두 잘 반영하기 위해 이미지 모든 면의 끝에 값이 0인 padding을 추가할 수 있다. 주로 filter size가 F x F 일 때 zero padding은 (F-1)/2만큼 사용된다.

Padding을 사용할 경우 output size에도 변화가 생긴다. 예를 들어 input size가 32 x 32 x 3, filter 10개의 크기가 5 x 5 x 3, stride=1, pad=2일 때의 output size는 (32 + 2 \* 2 – 5)/1+1 = 32이며, filter가 10개이므로 최종적인 size는 32 x 32 x 10이다.

위의 예시에서 filter 하나의 파라미터 개수는 5 x 5 x 3 +1(bias) = 76개이다. 따라서 모든 filter의 파라미터 개수는 760개이다.

- Convolutional Filter size

Convolutional layer에는 ReLU(nonlinear operation)이 들어가므로 convolutional layer가 여러 개 있다면 비선형적인 데이터를 더 잘 fitting할 수 있을 것이다. 따라서 receptive field가 같더라도 큰 크기의 layer를 한번 사용하는 것보다 작은 크기의 layer를 여러 번 사용하는 것이 더 expressive한 feature map을 만들 수 있는 방법이다. 또한 이러한 방법을 이용하면 파라미터 수도 줄일 수 있다.

- Spatial Dimension at Convolution Layer

Input

K: number of filters, F x F: filter size, S: stride, P: amount of zero padding

Output size where

, ,

Total number of parameter: weights, biases

- Backpropagation of convolution layer

- 1 x 1 convolution

: Dimension을 줄일 때, 또는 activation map 여러 개를 하나로 concatenate할 때 주로 사용

- Dilated convolution

: Filter와 중첩된 부분의 모든 데이터 중 얼마나 사용할 지에 대한 결정. Receptive field를 높이고 싶을 때 주로 사용한다.