

컨볼루션 신경망 개념 - 컨볼루션(Convolution), 풀링(Pooling)

컨볼루션 신경망(CNN)은 이미지 분야를 다루기에 최적화된 인공신경망 구조이다.

- 컨볼루션 신경망은 컨볼루션층(Convolution Layer)과 풀링층(Pooling Layer)로 구성되어 있다.

01. 컨볼루션층(Convolution Layer)

- 컨볼루션층은 컨볼루션 연산을 통해서 이미지의 특징을 추출해내는 역할을 한다.
- 원본 이미지에 커널을 이용해서 컨볼루션을 수행하면 커널의 종류에 따라 원본 이미지의 특징들(features)이 활성화 맵(Activation Map)으로 추출 됨.
- 어떤 커널을 사용하는가에 따라 원본 이미지의 다양한 특징을 추출할 수 있다

Edge Detection Kernel : 원본이미지의 모서리를 추출한다.

- 원본 이미지 자체를 사용하는 것보다 모서리만 추출된 특징 이미지를 이용하는 것이 더 효율적이다. (QA) 보통 아래가 일반적인 값일까?

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Sharpen Kernel : 원본 이미지를 좀 더 명확한 이미지로 바꾼다.

(QA) 보통 아래가 일반적인 값일까? 어떤 용도에 사용되는가?

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

Box Blur Kernel : 원본이미지를 흐리게(Blur) 만들 수 있다.

(QA) 보통 아래가 일반적인 값일까? 어떤 용도에 사용되는가?

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

02. 풀링(Pooling) : CNN을 구성하는 2번째 요소

- 풀링은 차원을 축소하는 연산을 수행한다.
- 최대값 풀링(Max Pooling), 평균값 풀링(Average Pooling), 최소값 풀링(Min Pooling) 이 있다.
- 최소값 풀링은 거의 사용되지 않는다.
- 최대값 풀링이 많이 사용된다.
- 어떤 차이가 있을까?

2-1 최대값 풀링

- 이미지의 X * X 크기 부분에서 가장 큰 값(Max Value) 하나를 추출해서 원본 이미지의 X x X개의 값을 1개의 값으로 축소
- 어떻게 가능한가?

동일한 원리로 평균값 풀링, 최소값 풀링이 가능하다.

풀링층의 장점

- (가) 이미지의 차원을 축소함으로써 필요한 연산량의 축소가 가능.
- (나) 이미지의 가장 강한 특징만을 추출하는 특징 선별 효과가 있다.
- (다) 모서리가 추출된 활성화 맵에서 최대값 풀링을 하면 가로 세로 차원이 축소됨.

컨볼루션 이후의 인풋 이미지의 결과 이미지

- 가로 길이가 Win이라면 출력 Wout은 다음과 같다.
- 세로 길이가 Hin이라면 출력 Hout은 다음과 같다.

$$W_{out} = \frac{(W_{in} - F + 2P)}{S} + 1$$

S : 스트라이드(컨볼루션 연산시 건너뛰는 정도)

-> 스트라이드(Stride)를 많이 잡으면 많이 축소되고,

-> 스트라이드(Stride)를 적게 잡으면 이미지가 조금 축소된다.

F : 필터의 크기

2P : 인풋 이미지의 상하좌우 모서리에 P만큼 0을 채워주는 제로패딩(zero padding)을 P만큼 적용

-> 좌우가 P만큼 갖기 때문에 2P만큼 가로 길이가 커진다.

$$H_{out} = \frac{(H_{in} - F + 2P)}{S} + 1$$

S : 스트라이드(컨볼루션 연산시 건너뛰는 정도)

-> 스트라이드(Stride)를 많이 잡으면 많이 축소되고,

-> 스트라이드(Stride)를 적게 잡으면 이미지가 조금 축소된다.

F : 필터의 크기

2P : 인풋 이미지의 상하좌우 모서리에 P만큼 0을 채워주는 제로패딩(zero padding)을 P만큼 적용

-> 좌우가 P만큼 갖기 때문에 2P만큼 가로 길이가 커진다.

컨볼루션 차원 출력 결과는 아래와 같다.

$$[W_{out}, H_{out}, K]$$

예제 MNIST

- [28 x 28 x 1] 이미지에 4 X 4 크기의 필터(F=4)에 스트라이드(Stride)가 S(=2)이고,
- 제로패딩을 적용하지 않은 (P=0)
- 필터개수 (K=64)를 가진 컨볼루션층을 적용하면

- $W_{out} = (28-4 + 2*0)/2 + 1 = 14$,
- $H_{out} = (24-4 + 2*0)/2 + 1 = 14$
- $K = 64$

출력 결과로 [14,14, 64], 즉 14 X 14 크기의 64개의 활성화맵이 추출될 것이다.

분류 문제를 위한 CNN의 경우, 컨볼루션과 풀링층을 거쳐서 추출된 활성화 맵은 마지막에 Flattening으로 펼친 이 후,

ANN 구조인 완전 연결층(Fully Connected Layer)의 인풋으로 들어가 Softmax 분류를 수행한다.

In []:

1