합성곱 신경망(CNN)

경제학과 2015231035 하지민

합성곱 신경망(CNN) 소개

합성곱 신경망

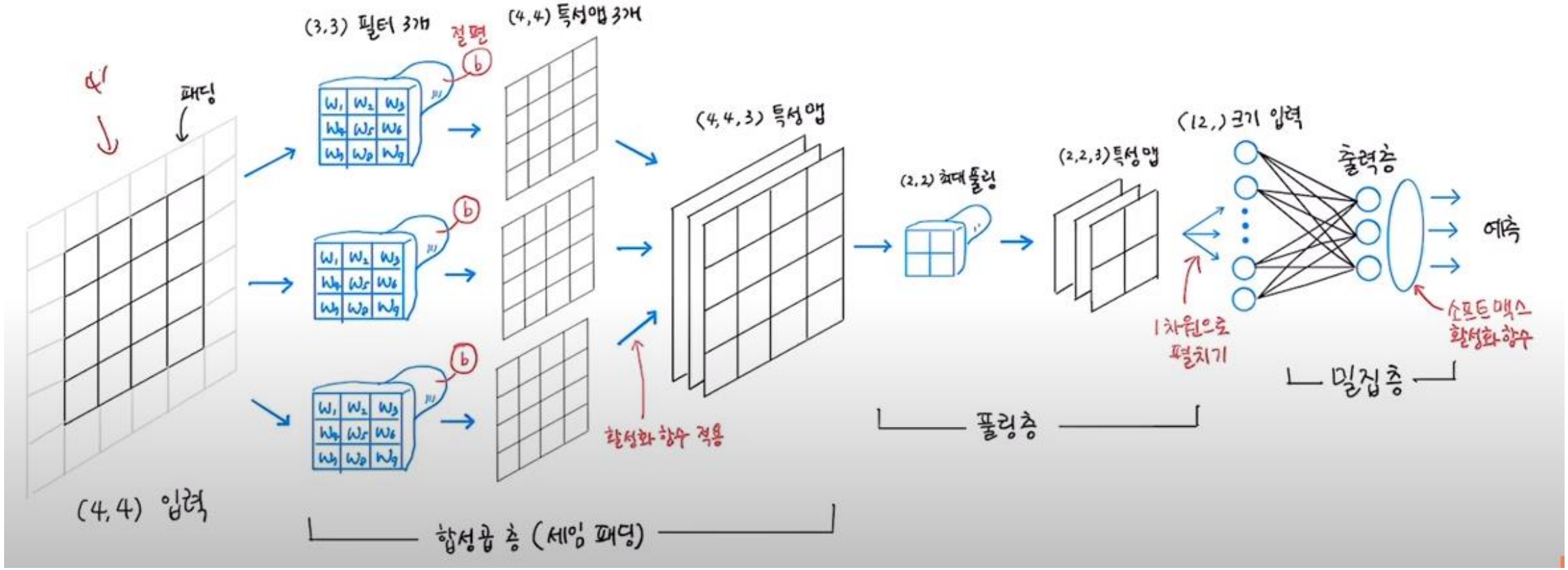
합성곱 신경망(CNN : Convolutional Neural Network)은 음성 인식이나 이미지 인식 에서 주로 사용되는 신경망의 한 종류
다차원 배열 데이터를 처리하도록 구성되어 있어, 컬러 이미지와 같은 다차원 배열 처리에 특화
이미지 인식 분야에서 딥러닝을 활용한 기법은 대부분 CNN을 기초로 함

완전연결 계층(밀집 신경망)의 문제점

'데이터 형상이 무시'
이미지 데이터의 경우 3차원의 형상
밀집 신경망에서 1차원 데이터로 펼치게 되면 이러한 공간적 구조에 대한 정보들이 사라짐

But,
합성곱 신경망(CNN)은 이미지를 하나의 데이터가 아닌, 여러 개로 분할하여 처리
이렇게 하면 이미지가 왜곡되더라도 이미지의 부분적 특성을 추출할 수 있어 올바른 성능을 낼 수 있음

합성곱 신경망의 구조


(3,3) 필터 3개
절편
(4,4) 특성맵 3개
(4,4,3) 특성맵
(2,2) 최대 풀링
(2,2,3) 특성맵
(12,) 크기 입력
출력층
예측
패딩
(4,4) 입력
활성화 함수 적용
1차원으로 펼치기
소프트맥스 활성화 함수
합성곱 층 (세임 패딩)
풀링층
밀집층

합성곱층

합성곱층은 합성곱 연산을 통해서 이미지의 특징을 추출하는 역할을 함

합성곱 신경망(CNN)에서 합성곱 계층(Convolutional Layer)의 입출력 데이터는 다차원이기에 이것을 특징 맵(Feature Map)이라고 함

합성곱 연산

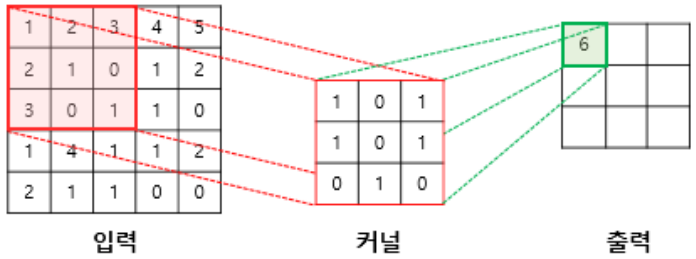
영어로 컨볼루션이라고 함
커널(kernel) 또는 필터(filter)라는 n x m 크기의 행렬로 높이(height) x 너비(width) 크기의 이미지를 처음부터 끝까지 겹치며 훑으면서 n x m크기의 겹치는 부분의 각 이미지와 커널의 원소의 값을 곱해서 모두 더한 값을 출력하는 것을 말함

합성곱층

합성곱 연산 순서

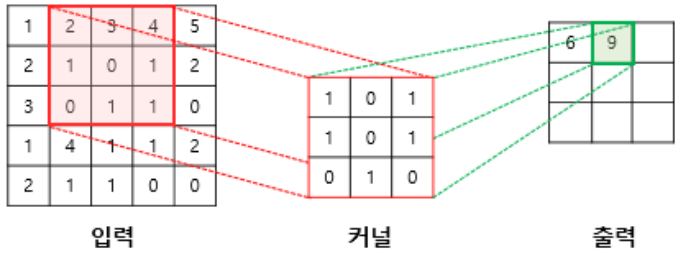
이미지는 가장 왼쪽 위부터 가장 오른쪽까지 순차적으로 훑음

1. 첫번째 스텝



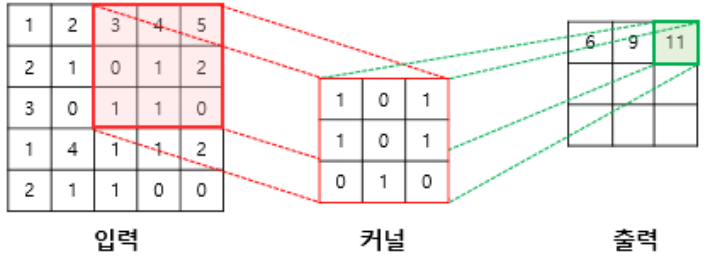
(1×1) + (2×0) + (3×1) + (2×1) + (1×0) + (0×1) + (3×0) + (0×1) + (1×0) = 6

2. 두번째 스텝



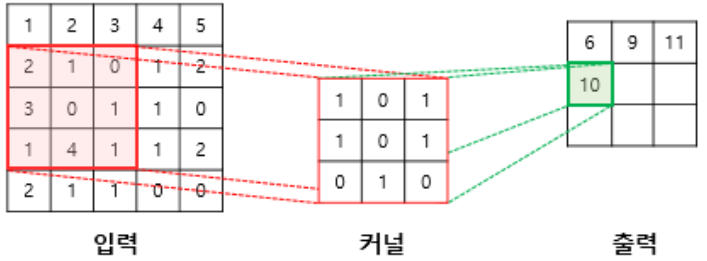
(2×1) + (3×0) + (4×1) + (1×1) + (0×0) + (1×1) + (0×0) + (1×1) + (1×0) = 9

3. 세번째 스텝



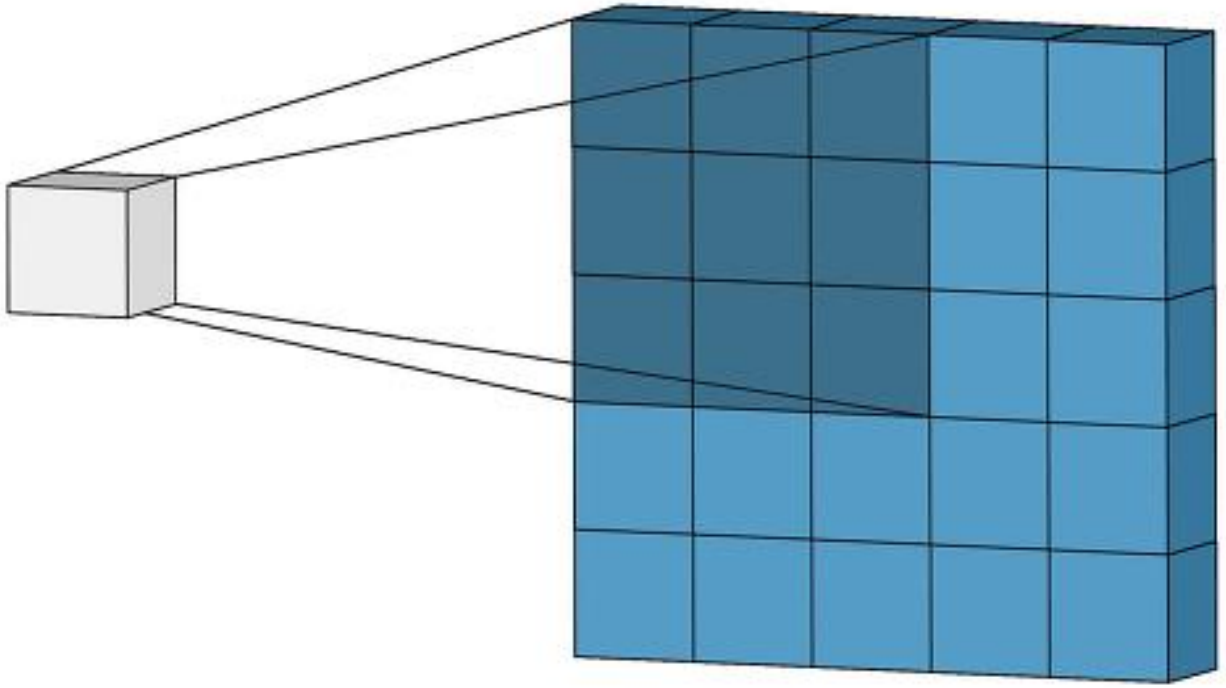
(3×1) + (4×0) + (5×1) + (0×1) + (1×0) + (2×1) + (1×0) + (1×1) + (0×0) = 11

4. 네번째 스텝



(2×1) + (1×0) + (0×1) + (3×1) + (0×0) + (1×1) + (1×0) + (4×1) + (1×0) = 10

합성곱층



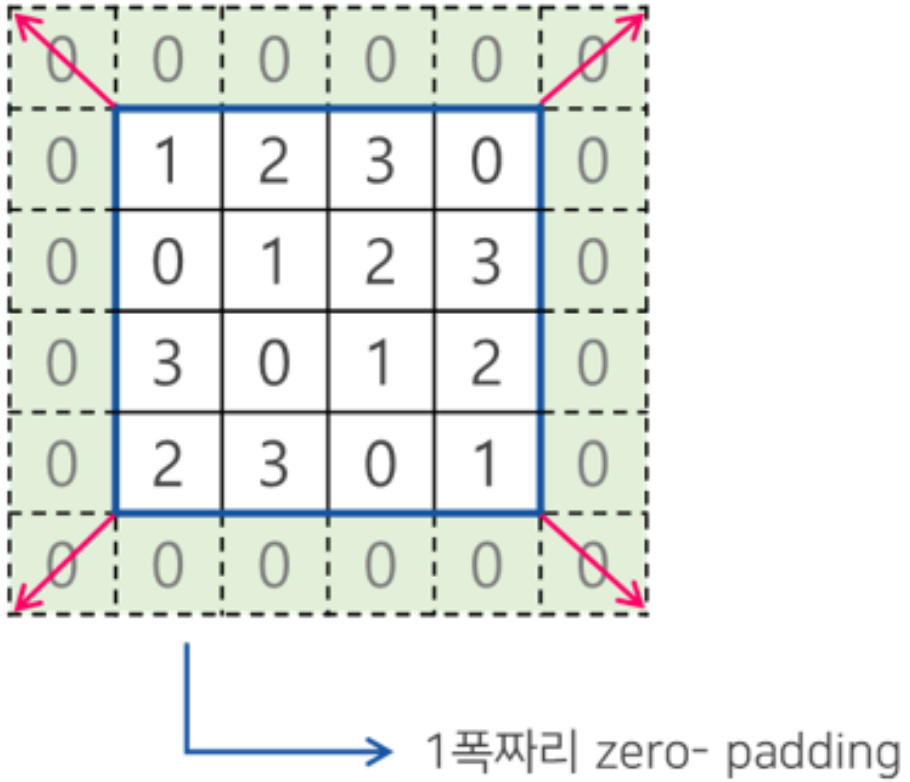
합성곱층

패딩

합성곱 연산을 수행하기 전에 입력 데이터 주변은 특정 값(0, 1 등)으로 채우기도 하는데, 이를 패딩이라 하며 합성곱 연산에서 자주 이용되는 기법

폭 1짜리 패딩이면 데이터 사방 1픽셀을 특정 값으로 채우는 것을 말함

주로 zero-padding을 사용함

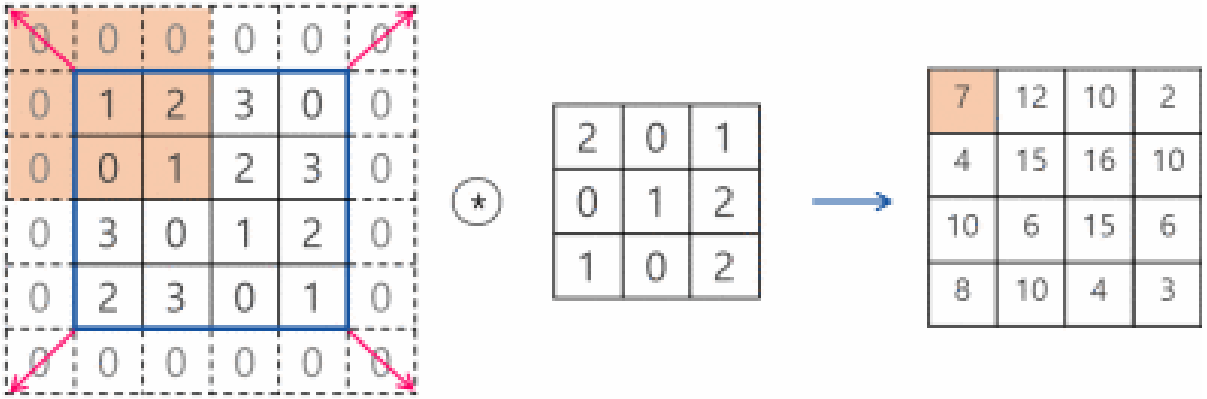


합성곱층

스트라이드

입력데이터에 필터를 적용할 때 이동할 간격을 조절하는 것, 즉 필터가 이동할 간격을 말함

스트라이드(Stride)는 보통 1과 같이 작은 값이 더 잘 작



합성곱층

출력 크기 계산

패딩과 스트라이드를 적용하고, 입력데이터와 필터의 크기가 주어졌을 때 출력 데이터의 크기를 구하는 식은 아래와 같다.

$$(OH, OW) = \left(\frac{H + 2P - FH}{S} + 1, \frac{W + 2P - FW}{S} + 1 \right)$$

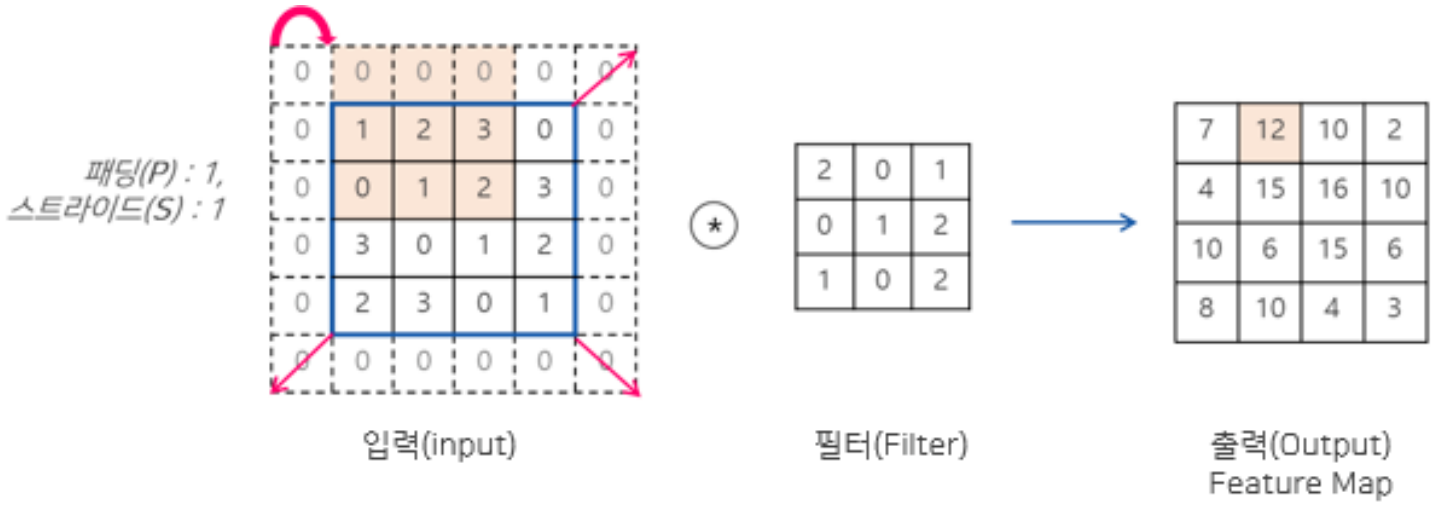
- (H, W) : 입력 크기 (input size)
- (FH, FW) : 필터 크기 (filter/kernel size)
- S : 스트라이드 (stride)
- P : 패딩 (padding)
- (OH, OW) : 출력 크기 (output size)

합성곱층

출력 크기 예시

아래의 그림은 패딩(padding) 1, 스트라이드(stride) 1 일 때의 출력데이터 크기를 구한 예제

출력크기가 정수가 아닌 경우에는 에러가 발생할 수 있는데, 보통 딥러닝 프레임워크에서는 반올림을 통해 에러 없이 작동

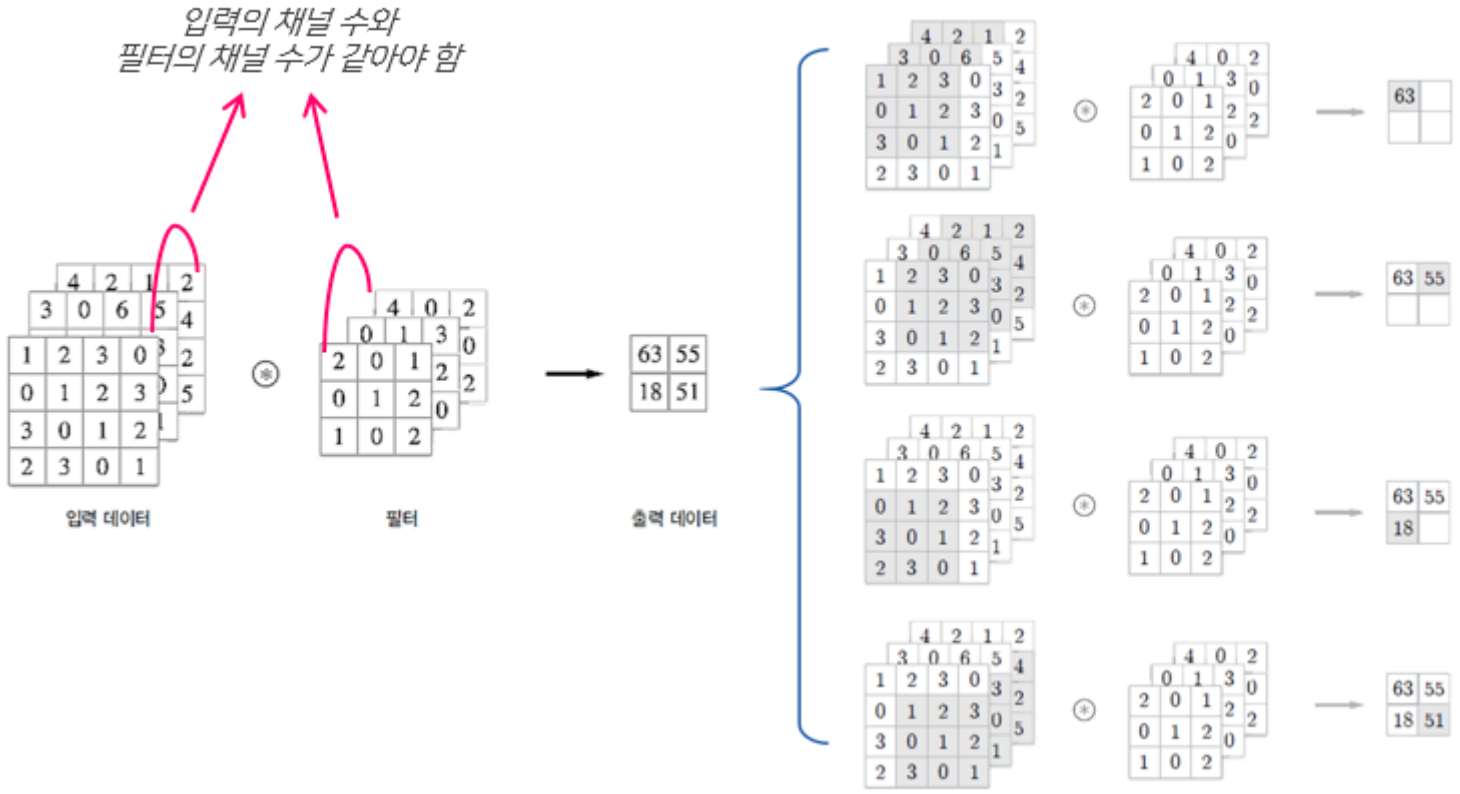


$$(\mathrm{OH}, \mathrm{OW}) = \left(\frac{4 + 2 \times 1 - 3}{1} + 1, \frac{4 + 2 \times 1 - 3}{1} + 1 \right) = (4, 4)$$

합성곱층

3차원 데이터 의 합성곱

3개의 채널을 가지는 이미지의 다음과 같이 합성곱 연산을 수행할 수 있는데, 여기서 주의해야할 점은
합성곱 연산을 수행할 때, 입력 데이터의 채널 수와 필터의 채널수가 같아야 함



풀링층

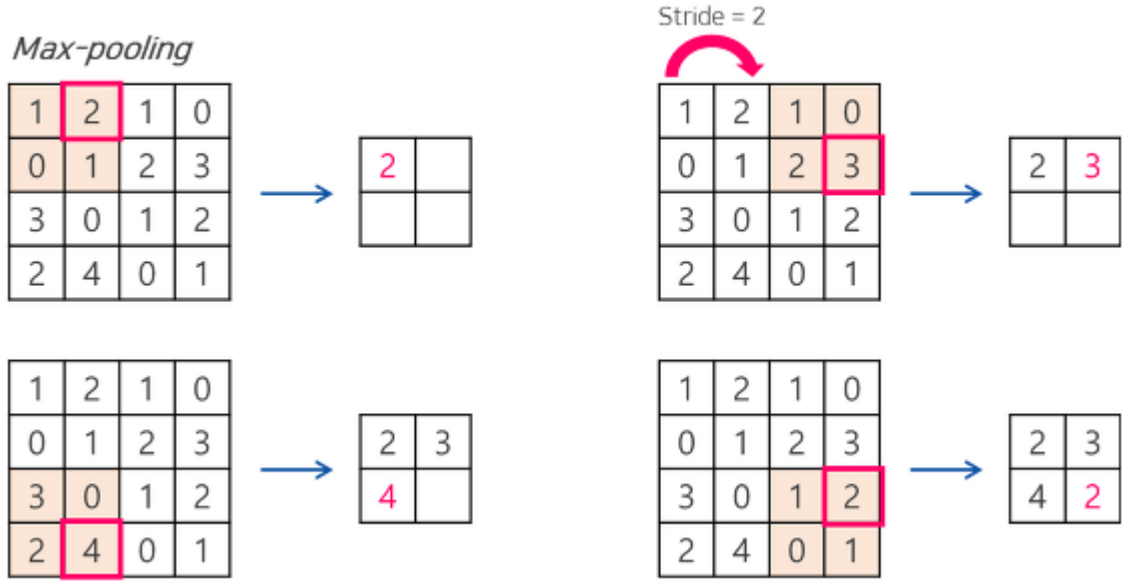
기술적 측면에서 풀링은 차례로 처리되는 데이터의 크기를 줄임

이 과정으로 모델의 전체 매개변수의 수를 크게 줄일 수 있음

풀링에는 Max-Pooling과 Average pooling 존재

Max-Pooling은 해당영역에서 최대값을 찾는 방법
Average-Pooling은 해당영역의 평균값을 계산하는 방법

이미지 인식 분야에서는 주로 Max-Pooling을 사용

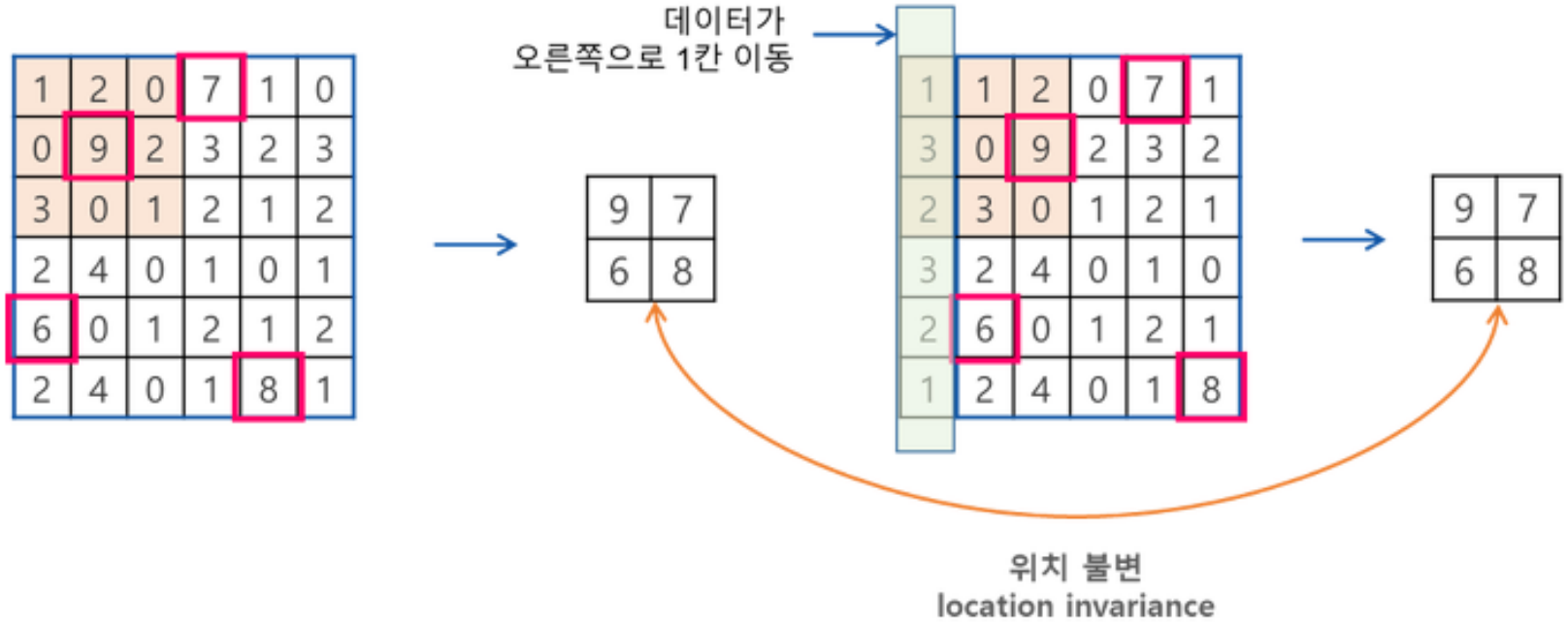


풀링층

이론적 측면은 계산된 특징이 이미지 내의 위치에 대한 변화에 영향을 덜 받기 때문

예를 들어 이미지의 우측 상단에서 눈을 찾는 특징은, 눈이 이미지의 중앙에 위치하더라도 크게 영향을 받지 않아야 함

그렇기 때문에 풀링을 이용하여 불변성(invariance)을 찾아내서 공간적 변화를 극복할 수 있음



참고 및 출처

유현준, 21.04.16, 딥러닝을 이용한 자연어 처리 입문(https://wikidocs.net/book/2155)

https://excelsior-cjh.tistory.com/180 [EXCELSIOR]