Machine Learning

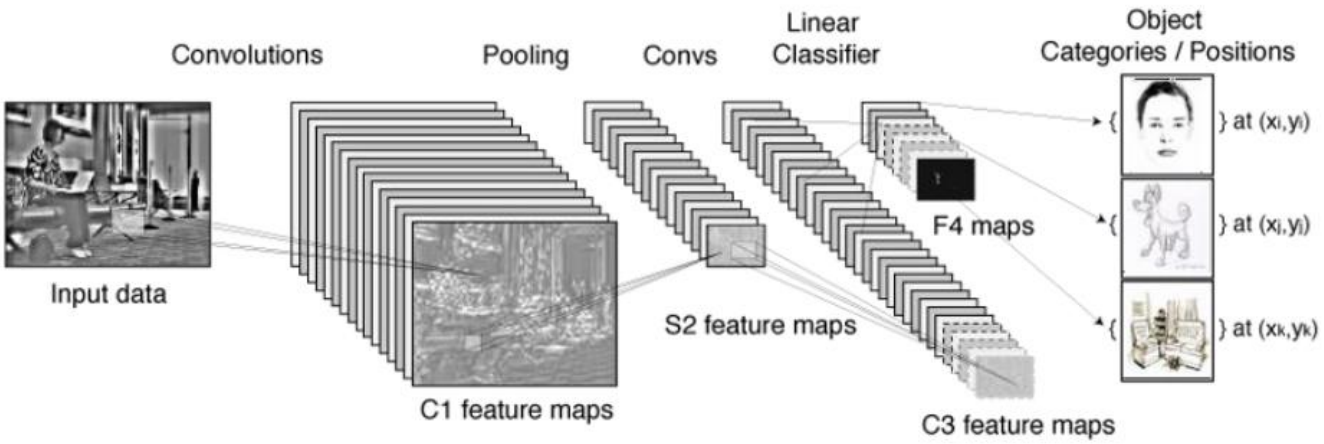
Chapter 7

합성곱 신경망(CNN)

7. 합성곱 신경망(CNN)

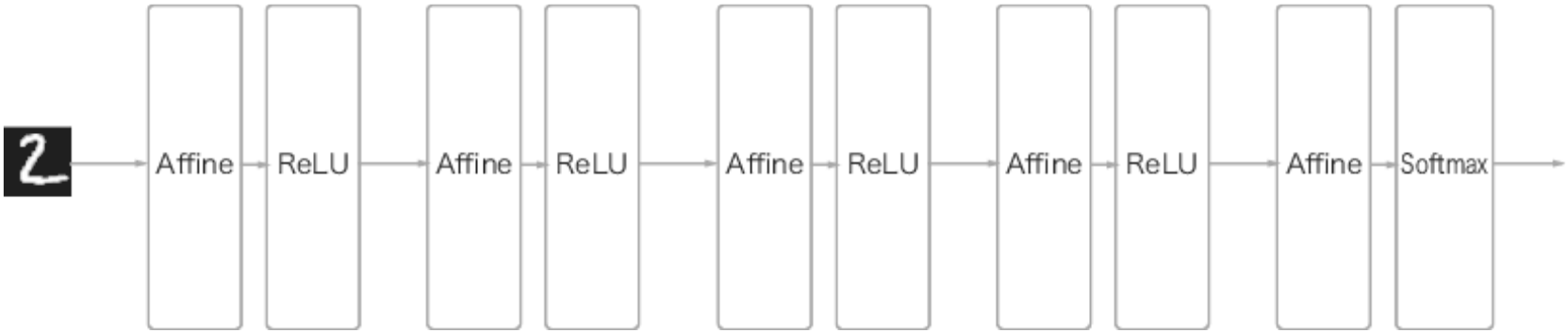
7.1 전체 구조

- **합성곱 신경망** convolutional neural network, CNN
  - 합성곱convolution 연산을 사용하는 ANNartificial neural network 의 한 종류
  - Convolution을 사용하면 3차원 데이터의 공간적 정보를 유지한 채 다음 레이어로 보낼 수 있다.
  - 대표적인 CNN으로는 LeNet(1998)과 AlexNet(2012)이 있으며, VGG, GoogLeNet, ResNet 등은 층을 더 깊게 쌓은 CNN 기반의 DNNdeep neural network, 심층 신경망 이다.

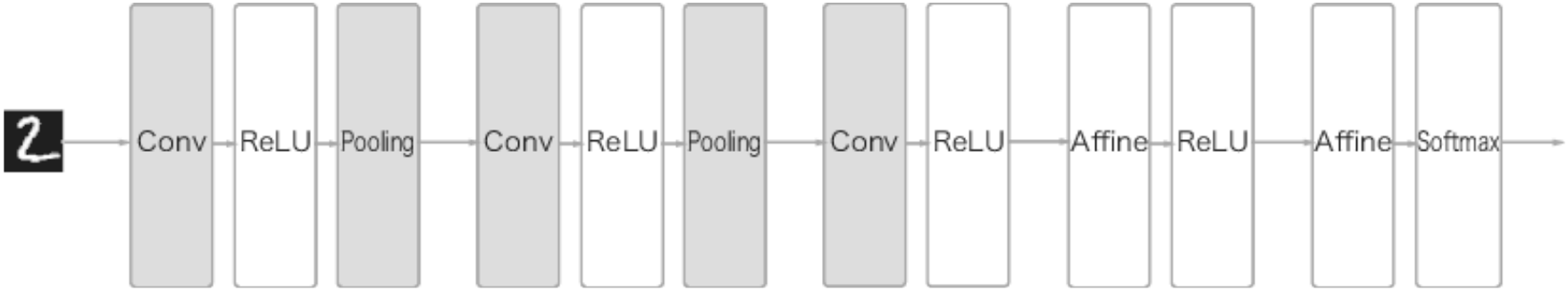


7.1 전체 구조

- 기존의 신경망 구조
  - 인접하는 계층의 모든 뉴런이 결합되어 있는 완전연결fully-connected 인 Affine 계층으로 구성되어 있다.

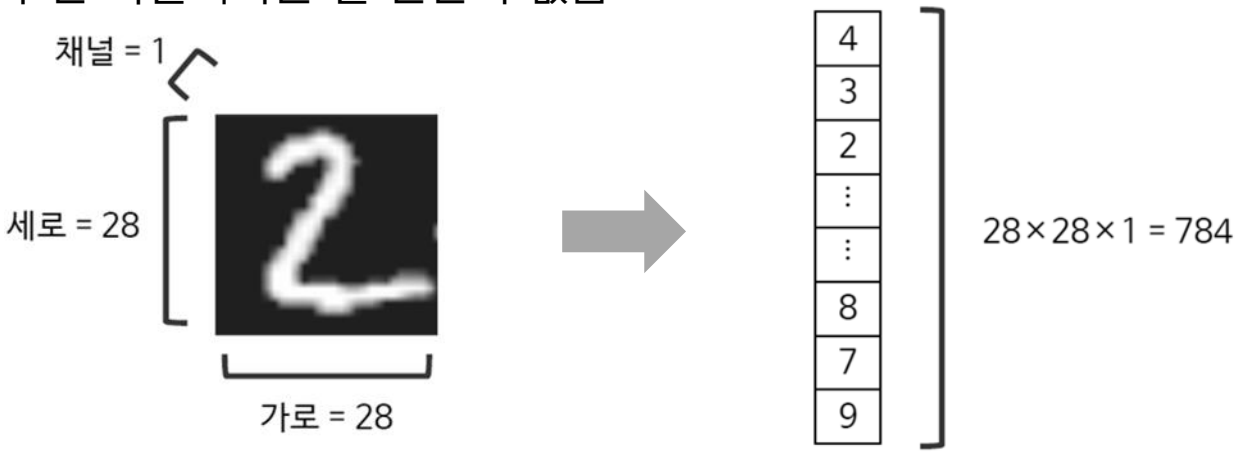


- CNN 구조
  - "합성곱 계층Conv"와 "풀링 계층Pooling"이 추가



7.2 합성곱 계층

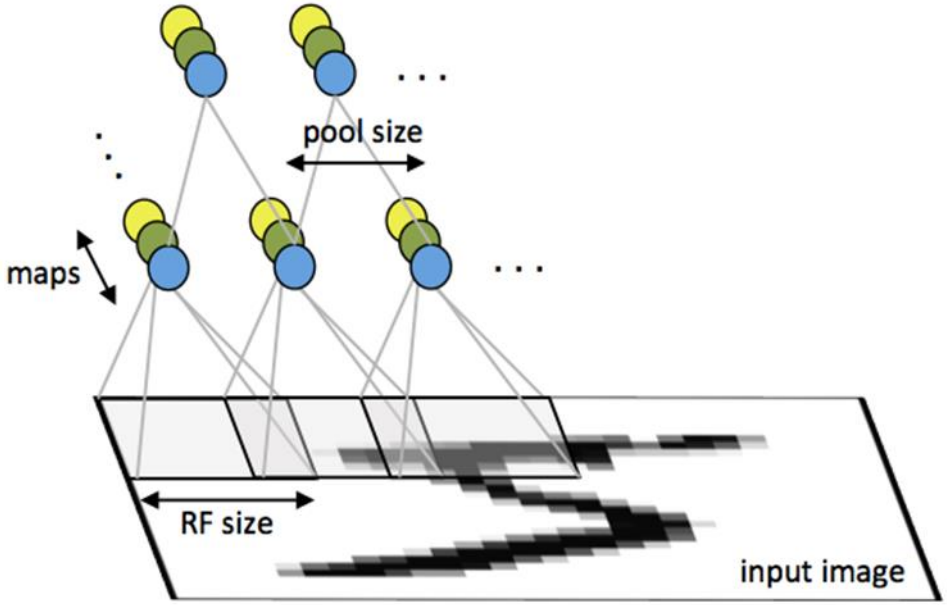
- **완전연결 계층의 문제점**
  - 데이터의 형상이 무시
    - 이미지는 세로 · 가로 · 채널(색상)로 구성된 3차원 데이터
    - 완전연결 계층에 입력할 때, 이미지를 1차원 데이터로 평탄화
    - 이미지에는 3차원 속에서 의미를 갖는 본질적인 패턴이 담겨 있다.
      * 공간적으로 가까운 픽셀은 값이 비슷함
      * RGB의 각 채널은 서로 밀접하게 관련됨
      * 거리가 먼 픽셀끼리는 별 연관이 없음



7.2 합성곱 계층

- 합성곱 계층은 형상을 유지한다.
- 이미지도 3차원 데이터로 입력받고, 다음 계층에도 3차원으로 전달

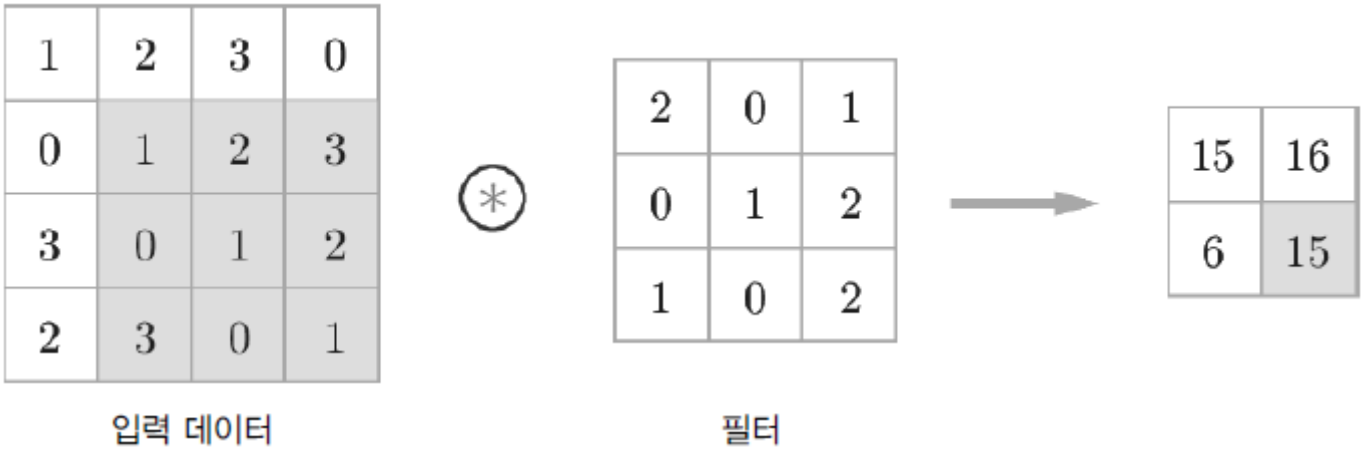
  ➡ 이미지처럼 형상을 가진 데이터를 제대로 이해할 것임
- 합성곱 계층의 입출력 데이터 : **특징 맵**feature map



7.2 합성곱 계층

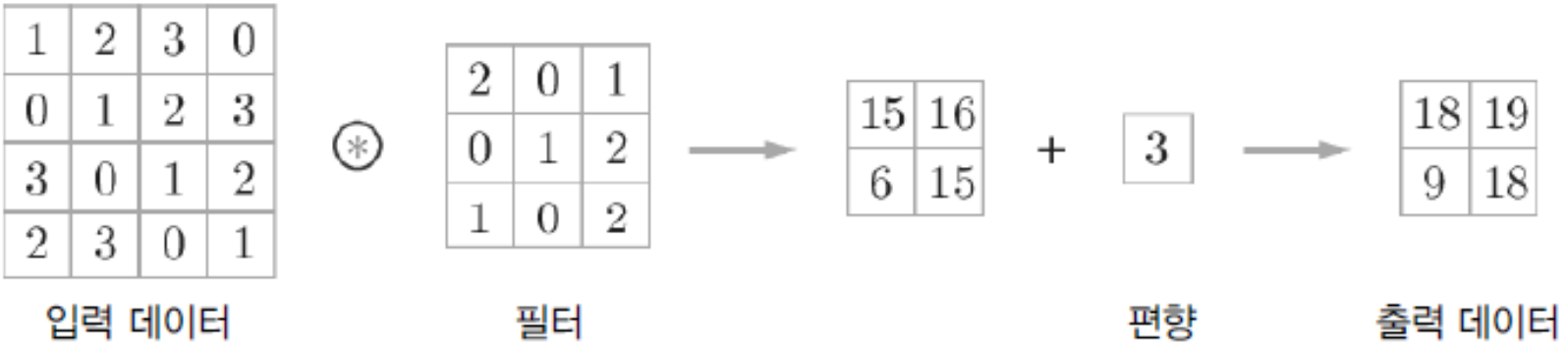
- **합성곱 연산**
  - 합성곱 연산은 입력 데이터에 **필터(커널)**를 적용
  - 필터의 **윈도우**window를 일정 간격으로 이동해가며 입력 데이터에 적용
    - 단일 곱셈-누산fused multiply-add, FMA

      : 입력과 필터에서 대응하는 원소끼리 곱한 후 그 총합을 구함

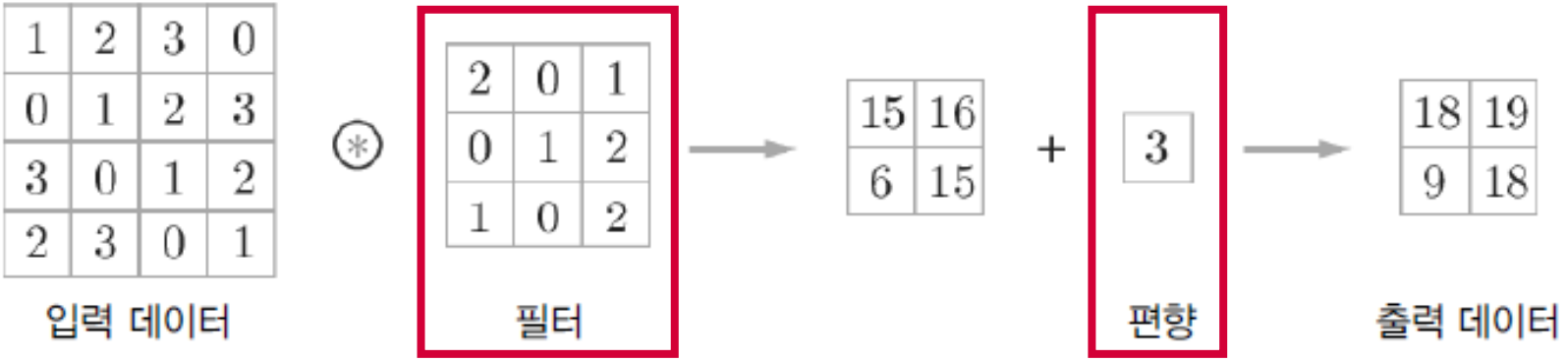


7.2 합성곱 계층

- **합성곱 연산**
  - 편향은 필터를 적용한 후의 데이터에 더해진다.



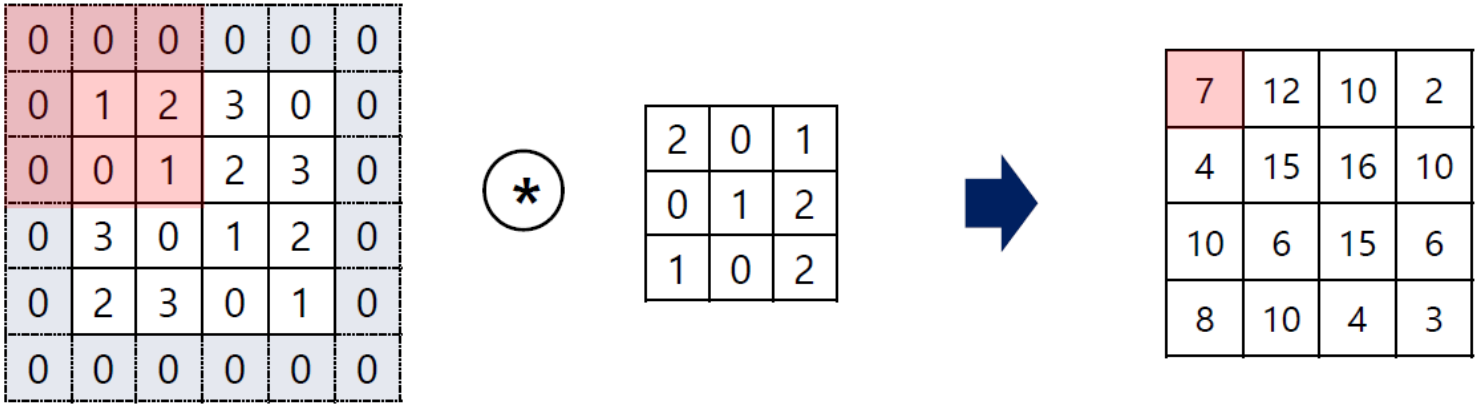
- **합성곱 연산의 매개변수**
  - 필터(가중치), 편향이 학습을 시킬 매개변수이다.



7.2 합성곱 계층

패딩 padding

- 합성곱 연산을 수행하기 전에 입력 데이터 주변을 특정 값(0, 1등)으로 채우는 것



Input Data
raw size - 4×4
after padding - 6×6

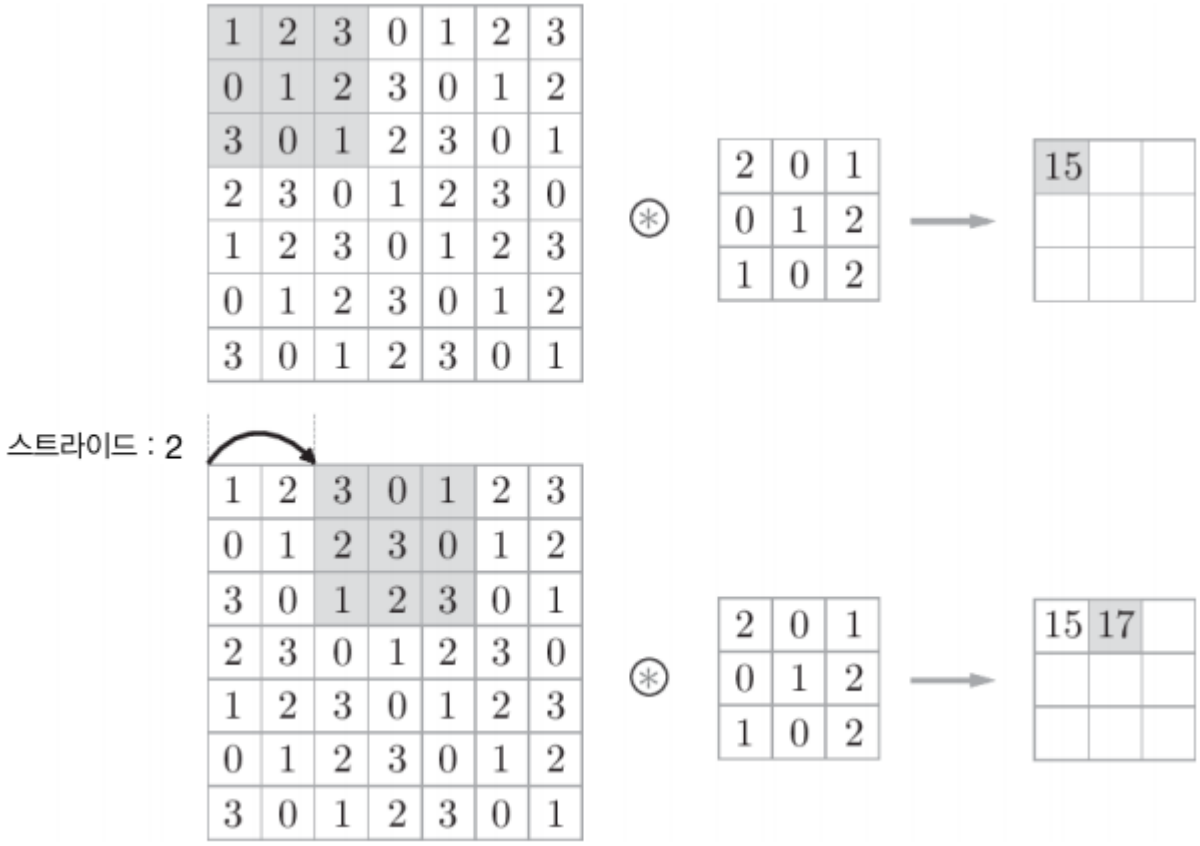
Filter
3×3

Output Data
4×4

패딩은 주로 출력 크기를 조정할 목적으로 사용한다. 합성곱 연산을 거칠 때마다 크기가 작아지면 어느 시점에서는 크기가 1이 되어 합성곱 연산을 적용할 수 없게 된다.
이를 막기 위해 패딩을 사용하여 입력 데이터의 공간적 크기를 고정한 채로 다음 계층에 전달할 수 있게 된다.

7.2 합성곱 계층

- 스트라이드 stride
  - 필터를 적용하는 위치의 간격
    - 스트라이드를 2로 하면 필터를 적용하는 윈도우가 두 칸씩 이동한다.



7.2 합성곱 계층

- 출력 크기 계산
  - 스트라이드를 크게 하면 출력크기는 작아진다.
  - 패딩을 크게 하면 출력크기가 커진다.

$$OH = \frac{H + 2P - FH}{S} + 1$$

$$OW = \frac{W + 2P - FW}{S} + 1$$

입력 크기 - (H, W)

필터 크기 - (FH, FW)

출력 크기 - (OH, OW)

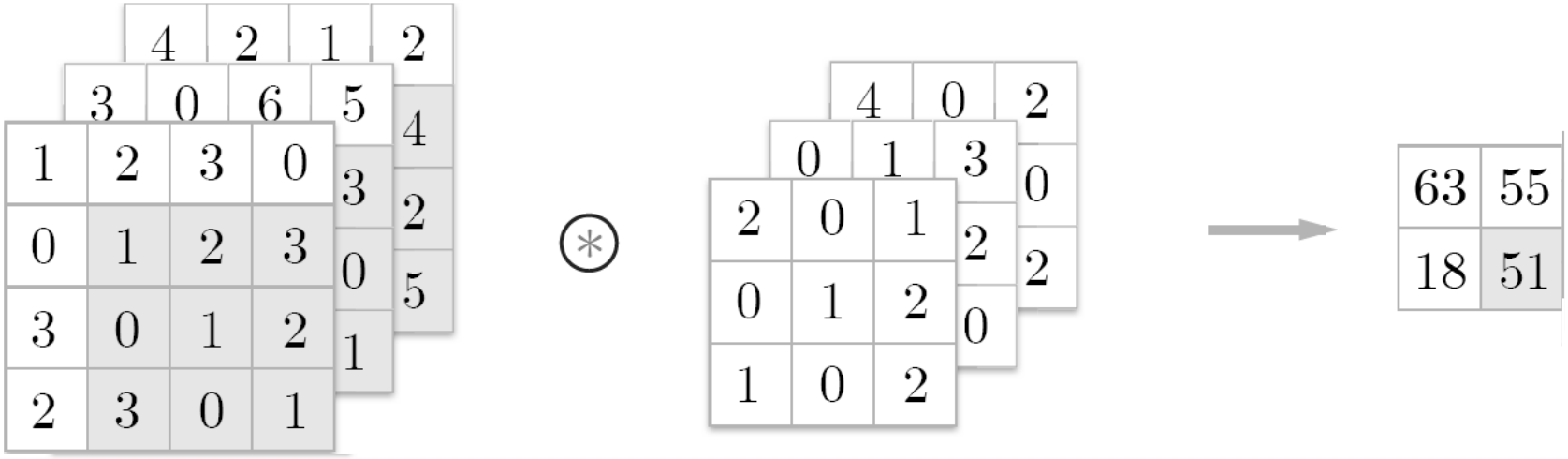
패딩 - P

스트라이드 - S

- 출력 크기(OH, OW)는 정수로 나눠떨어지는 값이어야 한다.
  * OH와 OW는 원소의 개수이므로
- 딥러닝 프레임워크 중에는 값이 딱 나눠떨어지지 않을 때는 가장 가까운 정수로 반올림 하는 경우도 있다.

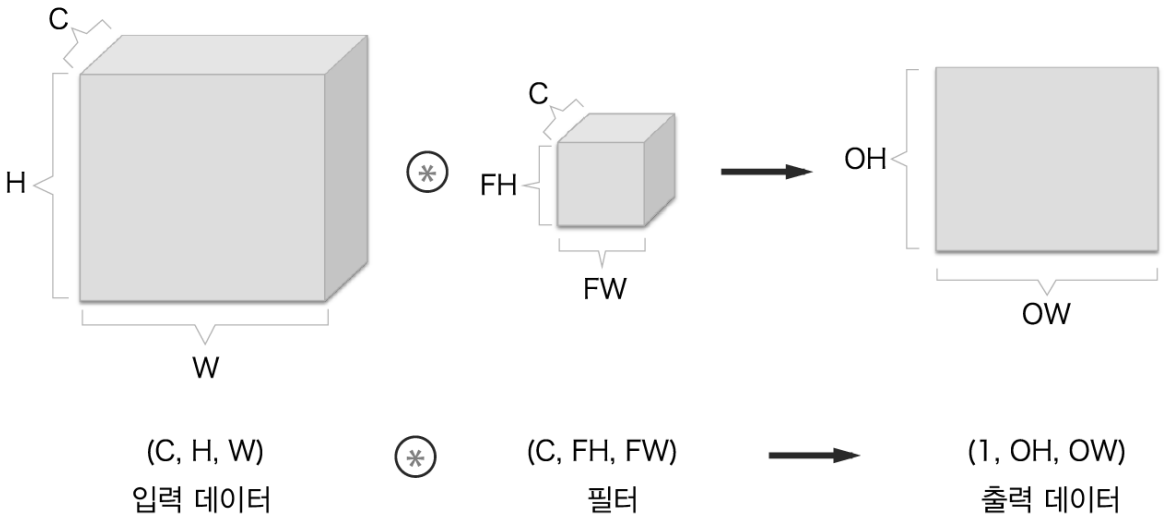
7.2 합성곱 계층

- 3차원 데이터의 합성곱 연산
  - 채널까지 고려한 3차원 데이터를 다루는 합성곱 연산
  - 입력 데이터와 필터의 합성곱 연산을 채널마다 수행하고, 그 결과를 모두 더해서 하나의 출력을 얻음
    - 입력 데이터와 필터의 채널 수가 같아야 함



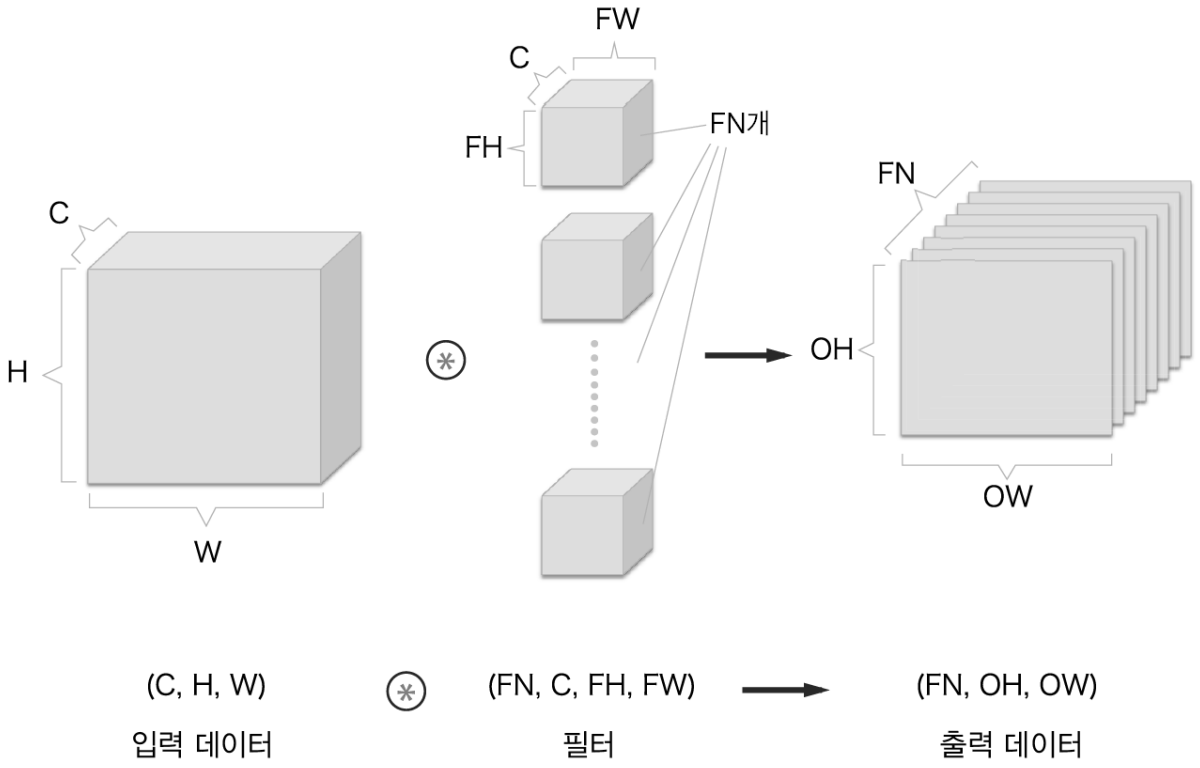
7.2 합성곱 계층

- 블록으로 생각하기
  - 하나의 필터를 사용한 합성곱 연산
    - 한 장의 특징 맵이 나오기까지의 과정
    - 필터의 형식은 (채널, 높이, 너비)로 나타냄



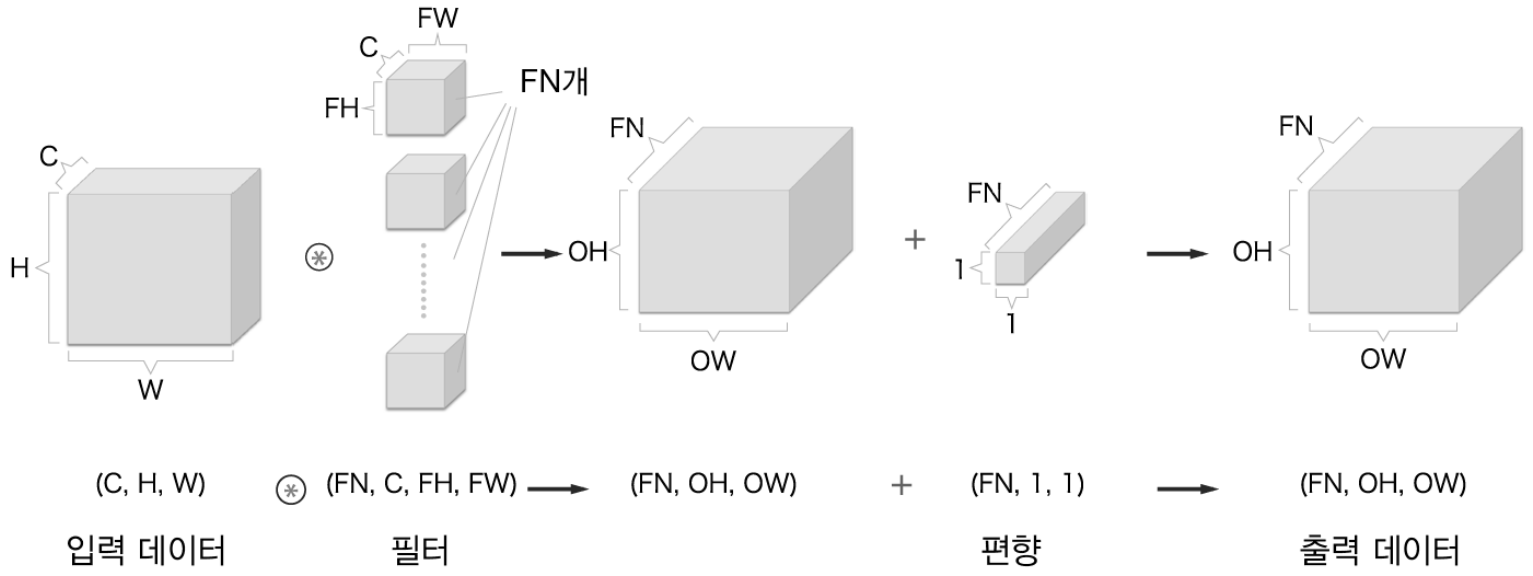
7.2 합성곱 계층

- 블록으로 생각하기
  - 여러 필터를 사용한 합성곱 연산
    - 여러 장의 특징 맵이 나오기까지의 과정
    - 필터의 형식은 (출력 채널 수, 입력 채널 수, 높이, 너비)로 나타냄



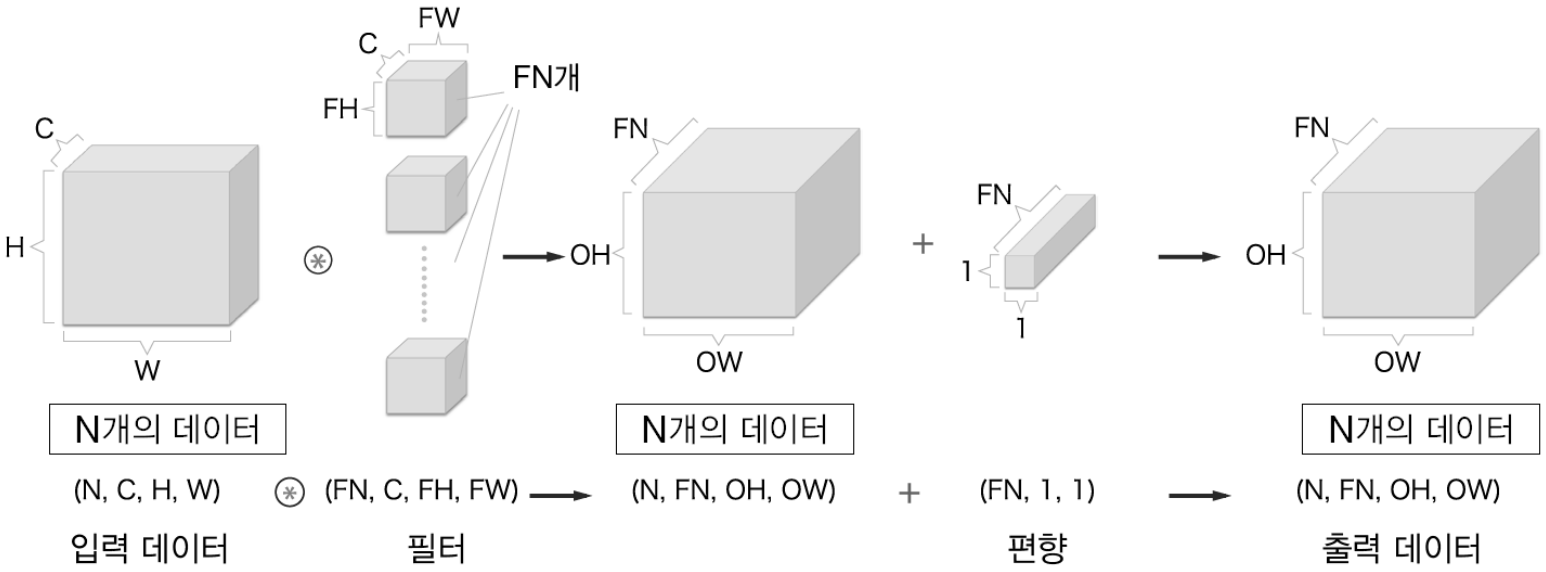
7.2 합성곱 계층

- 블록으로 생각하기
  - 합성곱 연산의 처리 흐름 + 편향
    - 편향은 채널 하나에 값 하나씩으로 구성
    - 각 편향이 필터의 출력인 (FN, OH, OW) 블록의 대응 원소에 더해짐



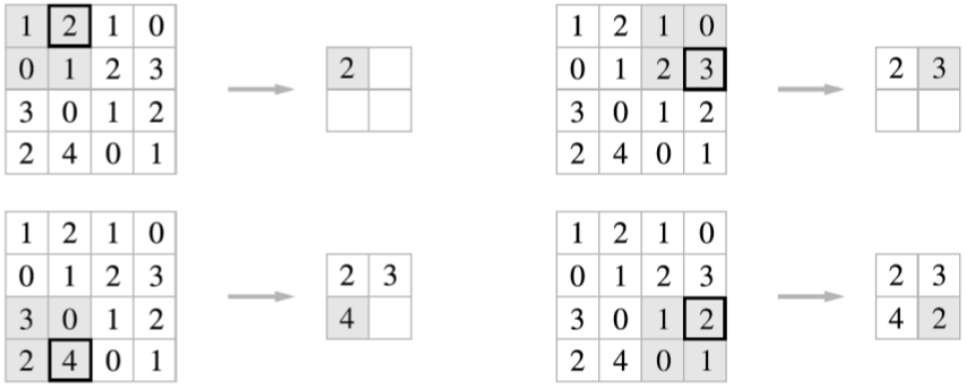
7.2 합성곱 계층

- **배치 처리**
  - 데이터 N개에 대한 합성곱 연산
    - 각 계층을 흐르는 데이터의 차원을 하나 늘려
      4차원 데이터(데이터 수, 채널 수, 높이, 너비)로 저장
    - 4차원 데이터가 흐를 때마다 데이터 N개에 대한 연산이 이루어짐
      → N회 분의 처리를 한번에 수행하는 것



7.3 풀링 계층

- **풀링** pooling
  - 세로 · 가로 방향의 공간을 줄이는 연산(sub-sampling)
  - 보통 풀링의 윈도우 크기와 스트라이드는 같은 값으로 설정

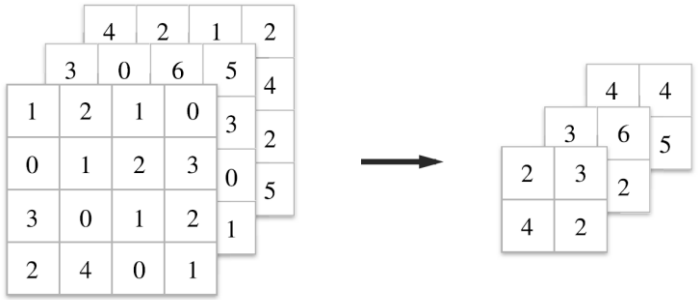


  - 최대 풀링max pooling : 대상 영역의 최대값
  - 평균 풀링average pooling : 대상 영역의 평균
  - 이미지 인식 분야에서는 주로 최대 풀링을 사용

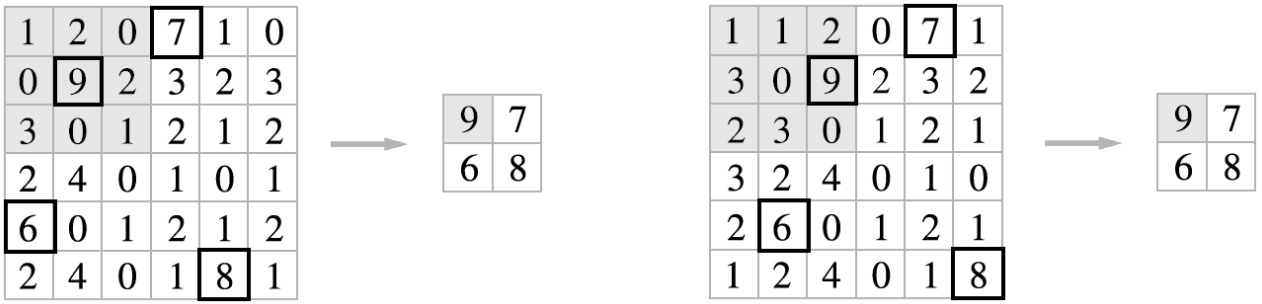
- **풀링 계층의 특징**
  - 학습해야 할 매개변수가 없다
    - 대상 영역에서 최대값이나 평균을 취하는 명확한 처리이므로 특별히 학습할 것이 없음

7.3 풀링 계층

- 채널 수가 변하지 않는다
  - 채널마다 독립적으로 계산하기 때문에
    입력 데이터의 채널 수 그대로 출력 데이터로 내보냄



- 입력의 변화에 영향을 적게 받는다(강건하다)
  - 입력 데이터가 조금 변해도 풀링의 결과는 잘 변하지 않음

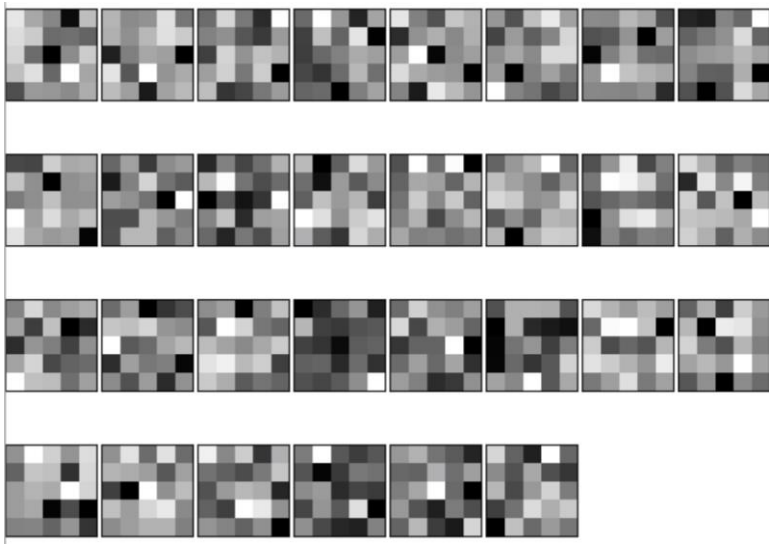


데이터가 오른쪽으로 1칸씩 이동

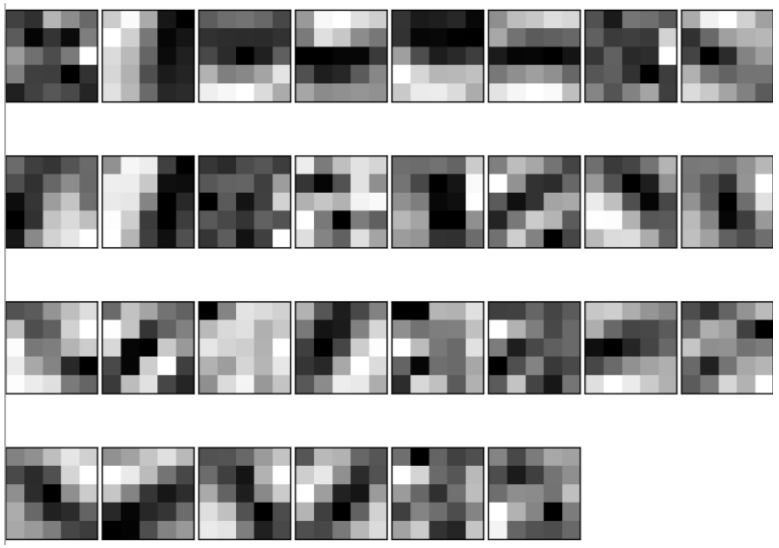
7.6 CNN 시각화하기

- **1번째 층의 가중치 시각화하기**
  - 학습 전 필터는 무작위로 초기화되어 흑백의 정도에 규칙성이 없음
  - 학습을 마친 필터는 규칙성 있는 이미지가 됨
    - 에지edge(색상이 바뀐 경계선)
    - 블롭blob(국소적으로 덩어리진 영역)

학습 전

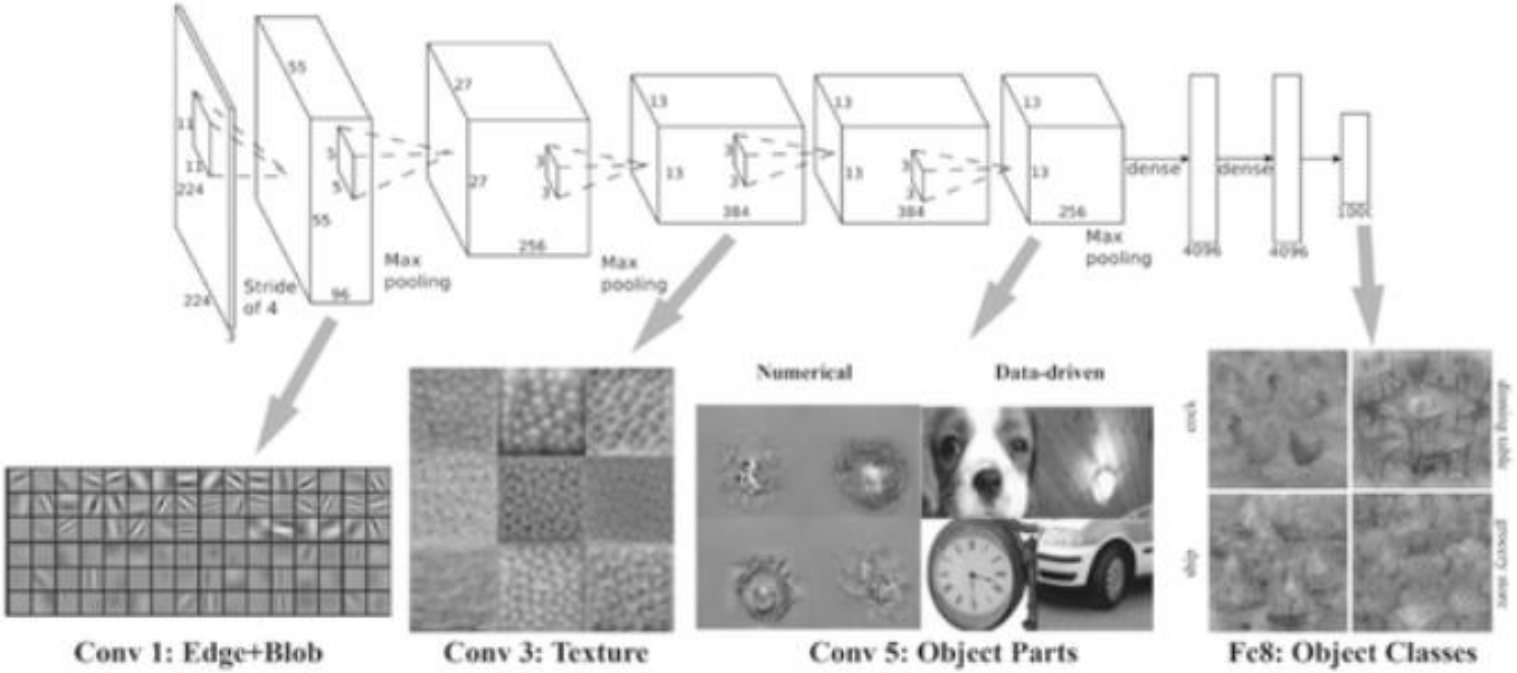


학습 후



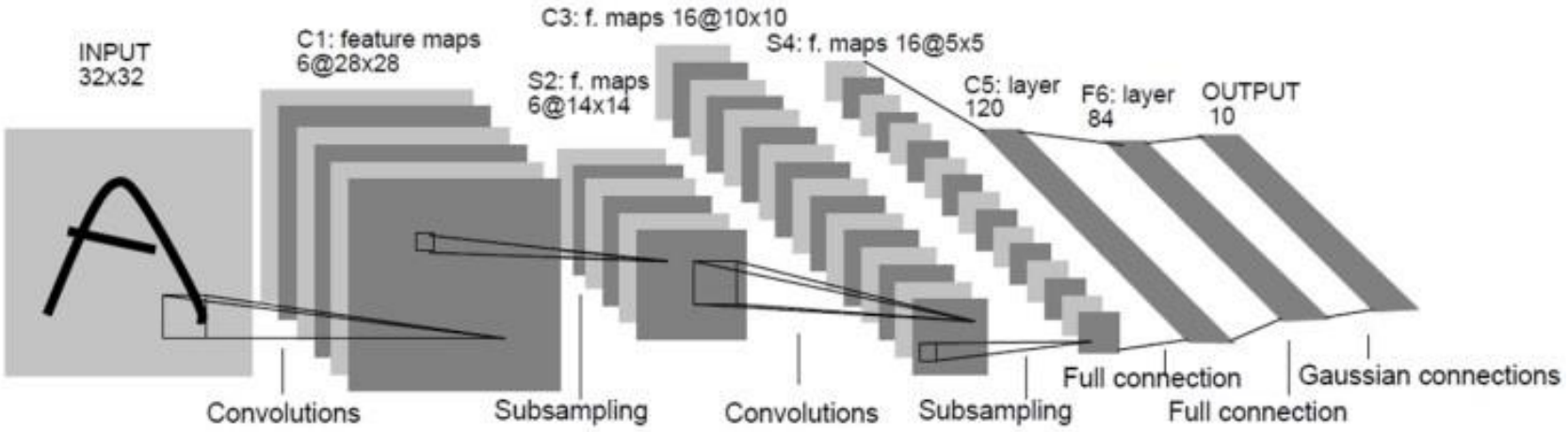
7.6 CNN 시각화하기

- **층 깊이에 따른 추출 정보 변화**
  - 계층이 깊어질수록 추출되는 정보(강하게 반응하는 뉴런)는 추상화
    → 층이 깊어지면서 고급 정보(사물의 의미)를 이해
    - 1번째 층 : 에지와 블롭
    - 3번째 층 : 텍스쳐
    - 5번째 층 : 사물의 일부
    - 마지막 FC층 : 사물의 클래스



7.7 대표적인 CNN

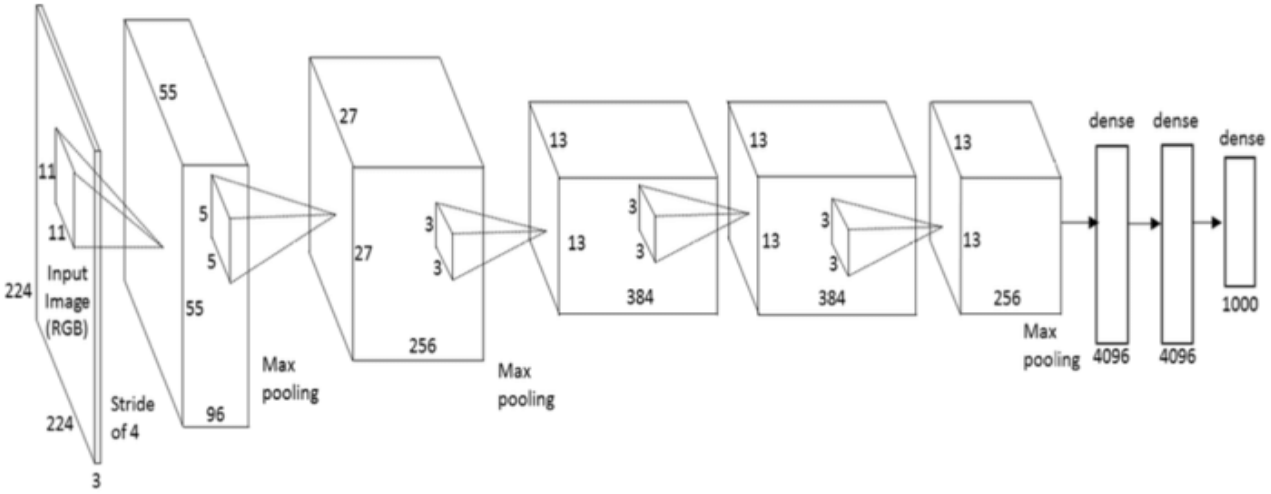
- LeNet
  - 손글씨 숫자를 인식하는 네트워크, 1998년에 제안된 첫 CNN
    - 활성화 함수로 시그모이드 함수를 사용
      - * 현재는 주로 ReLU를 사용
    - 서브샘플링을 하여 중간 데이터의 크기가 작아짐
      - * 현재는 Max Pooling이 주류



7.7 대표적인 CNN

- AlexNet
  - 2012년에 발표, 딥러닝 열풍을 일으킴
    * ILSVRC(ImageNet Large Scale Visual Recognition Challenge) – 2012 에서 우승
    - 활성화 함수로 ReLU를 이용
    - LRN Local Response Normalization 이라는 국소적 정규화 계층을 이용
    - 드롭아웃을 사용

      + Overlapped pooling, 2개의 GPU 사용



7.8 정리

이번 장에서 배운 내용

- CNN은 지금까지의 완전연결 계층 네트워크에 합성곱 계층과 풀링 계층을 새로 추가한다.
- 합성곱 계층과 풀링 계층은 im2col(이미지를 행렬로 전개하는 함수)을 이용하면 간단하고 효율적으로 구현할 수 있다.
- CNN을 시각화해보면 계층이 깊어질수록 고급 정보가 추출되는 모습을 확인할 수 있다.
- 대표적인 CNN에는 LeNet과 AlexNet이 있다.
- 딥러닝의 발전에는 빅데이터와 GPU가 크게 기여했다.