

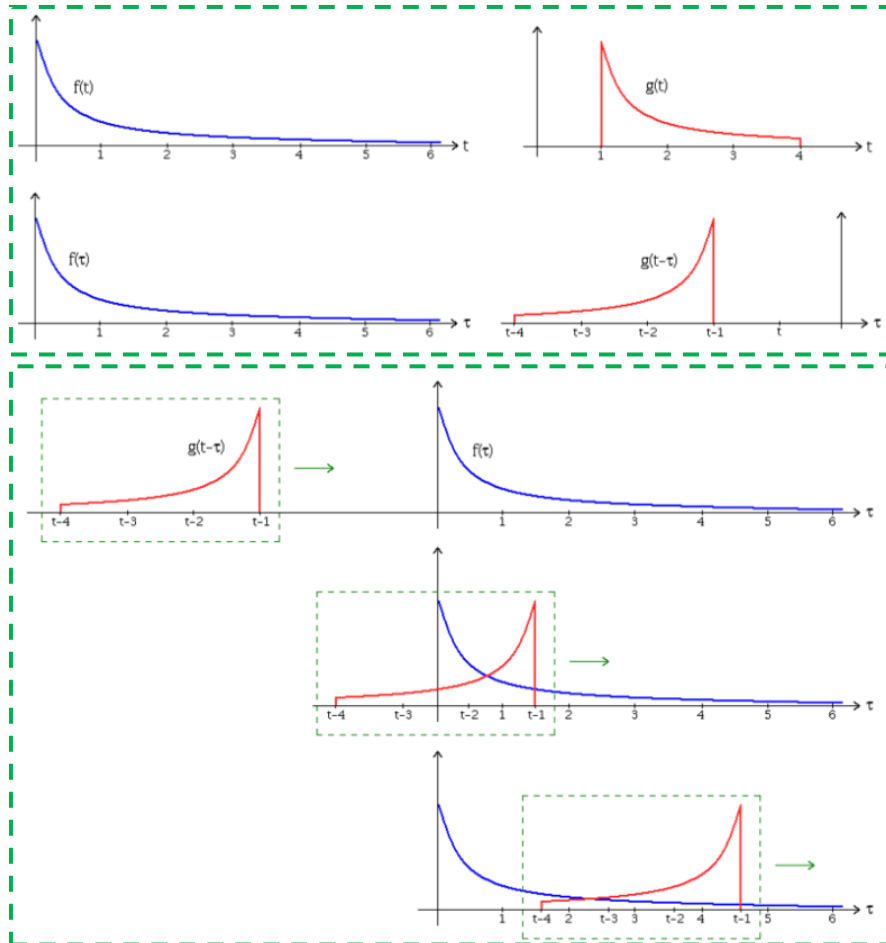
# CNN 개요



- **CNN (Convolutional Neural Network) 모델**

- 최근의 영상 인식, 처리 분야에 있어서 가장 기본이 되는 모델
- 눈과 뇌에서 처리되는 신경과학적 시각처리 방식에서 고안한 모델
- 신경망 동작을 처리하기 위하여 합성곱 연산을 이용함
  - 합성곱: 두 개의 함수가 있을 때, 둘 중에서 하나의 함수를 반전, 이동(전이)시킨 후, 두 함수를 곱한(결합한) 결과를 적분하여 그 파형(그래프)을 얻는 연산 방법
  - CNN 모델에서는 합성곱 연산을 기반으로 각 영역(픽셀)이 서로 얼마나 일치하는지 계산하여 그 계산 결과를 활용함

## • Convolution (합성곱)



반전

결과 예시

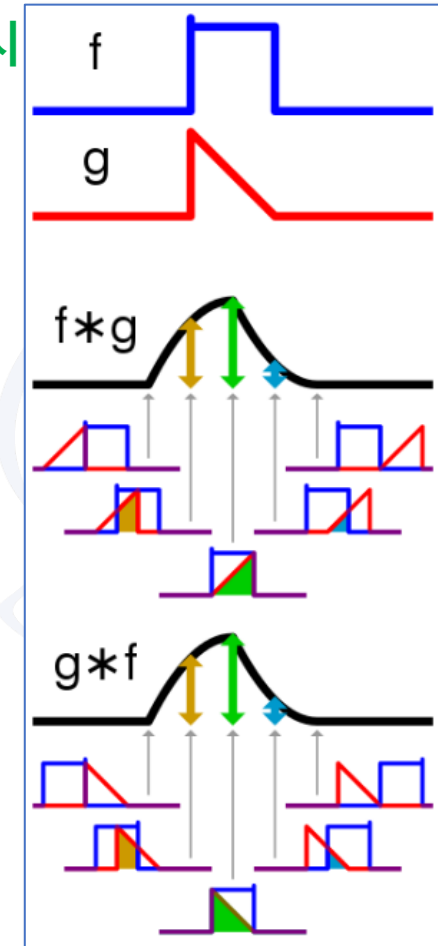
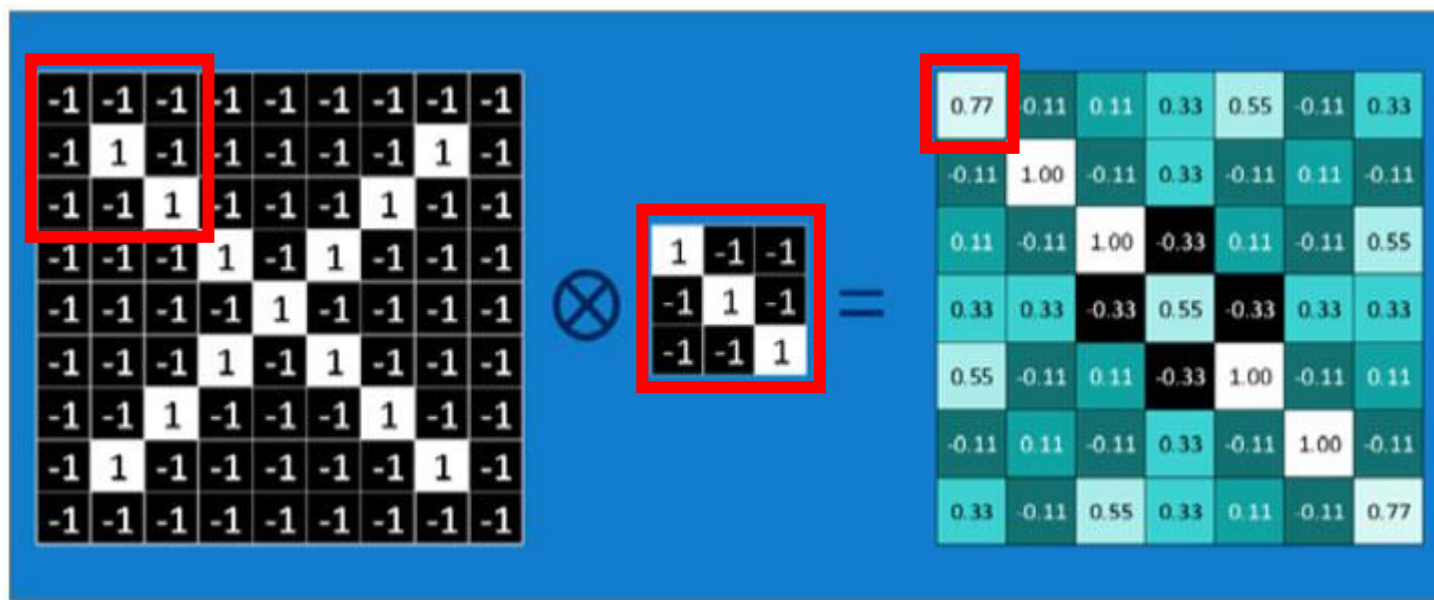


그림 출처: 위키피디아

## • Convolution (합성곱)

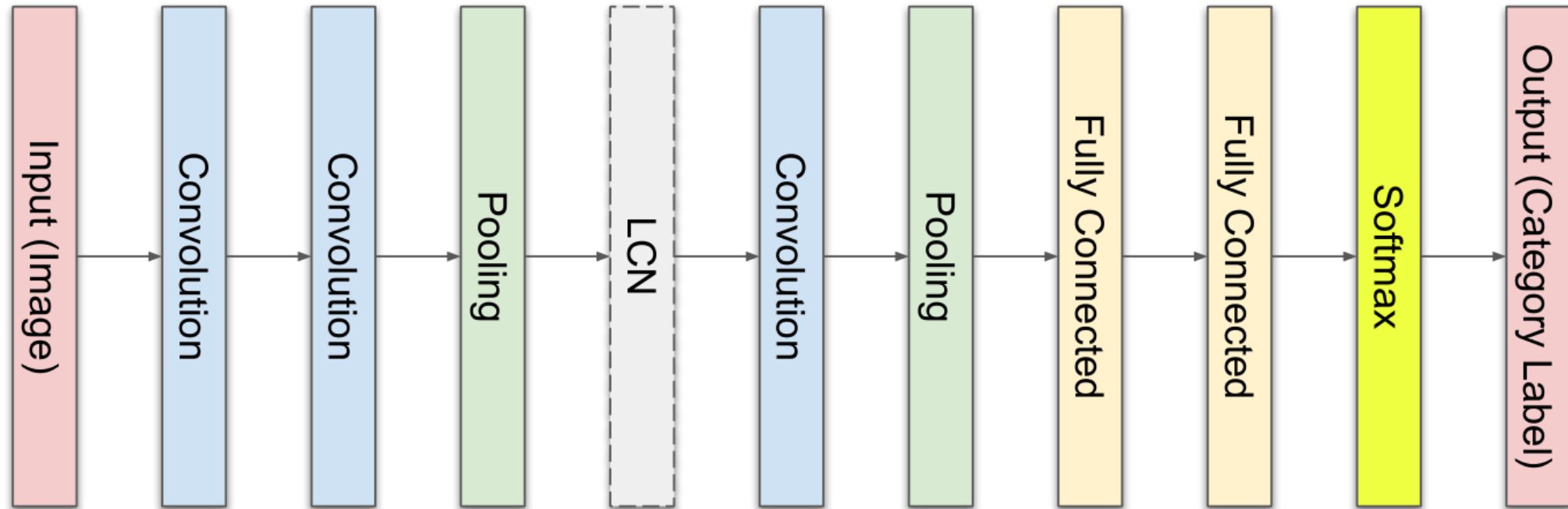
- 합성곱 신경망에서는 “하나의 함수가 다른 함수와 얼마나 일치하는가?”의 의미로 사용
- 하나의 필터(커널)에 대하여 이미지의 각 부분들이 필터와 얼마나 일치하는지 계산



- ① 각 요소를 1:1 대응으로 곱한 후 총합 계산  
 $(-1 + 1 + 1) + (1 + 1 + 1) + (1 + 1 + 1) = 7$
- ② 총합을 필터의 요소 개수(9개)를 기준으로 평균 계산  
 $7 / 9 = 0.777777..$

그림 출처: "파이토치 첫걸음" (최건호 저, 한빛미디어)

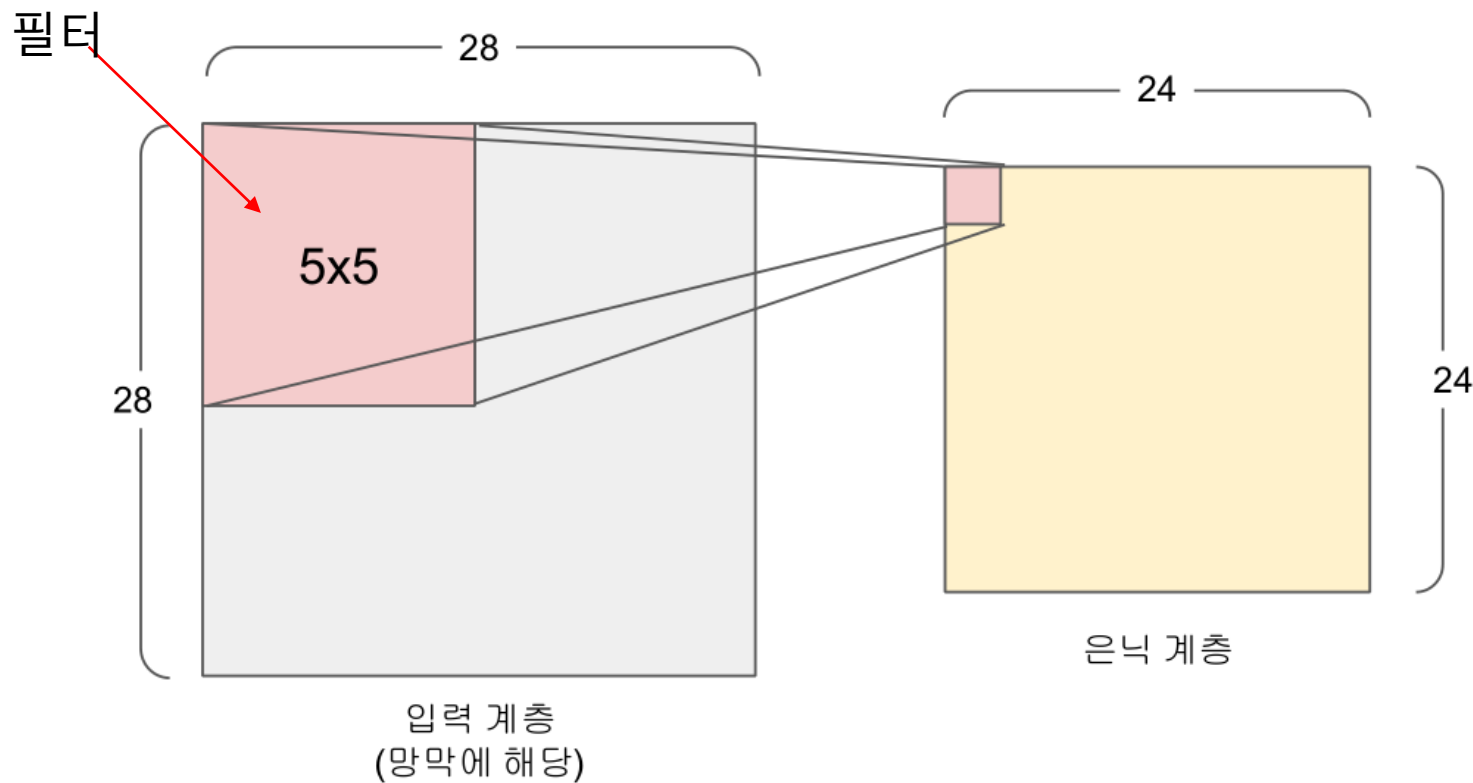
- CNN (Convolutional Neural Network) 모델
  - Convolution(합성곱) 을 이용하여 가중치 수 줄임 → 연산량 감소와 효과적인 이미지 처리 수행
  - Convolution Filter 를 이용하여 신경망 동작 수행
  - Convolution Layer 와 Fully Connected Layer 를 중심으로 영상정보 처리
    - Convolution Layer : 특징점을 효과적으로 찾는 데 활용
    - Fully Connected Layer : 발견한 특징점을 기반으로 이미지를 분류하는 데 활용



- Convolution Layer - Pooling Layer의 반복
- 반복 방식은 사용자가 임의로 결정 (예, Convolution Layer만 여러 층을 가진 후 마지막에 Pooling Layer가 오는 경우도 있음)
- 국소/지역 콘트라스트 정규화 (Local Contrast Normalization, LCN) 층을 매치하는 경우도 있음

# Convolution Layer (합성곱 층)

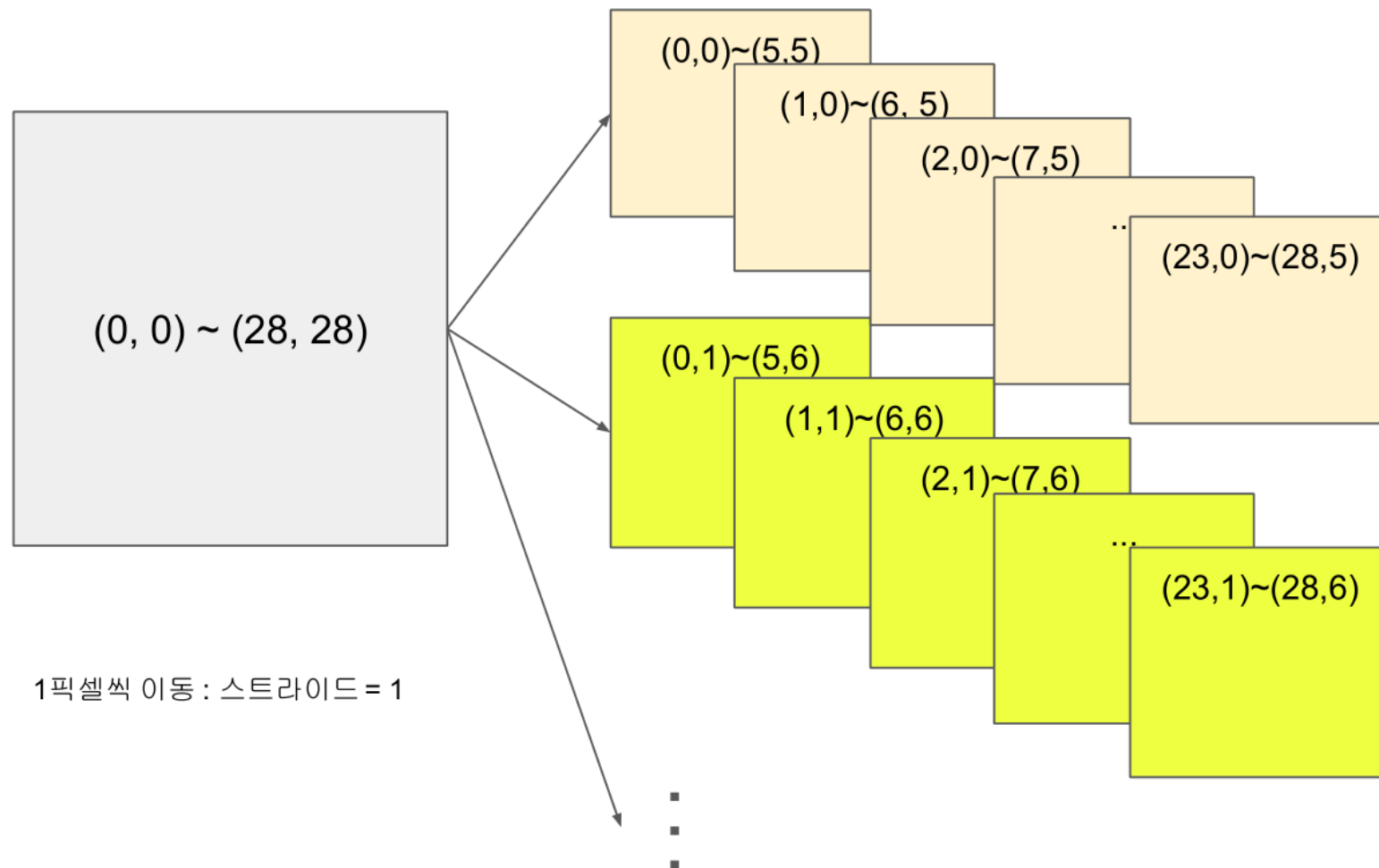
- 목적: 테두리, 선, 색 등 이미지의 시각적 특징이나 특성 감지



- 입력 계층과 은닉계층이 1:1 매핑되지 않음
- 시각피질의 수용야에 해당하는 윈도우는 정해진 크기만큼 좌표를 이동하며 입력계층의 데이터를 탐색함
- 픽셀 차이를 입력 계층과 겹쳐가며 이동하는데 이 간격을 스트라이드라고 함
- 필터의 크기(윈도우)가  $5 \times 5 = 25$  인 경우는 25개의 픽셀이 하나의 뉴런에 대응됨

# Convolution Layer (합성곱 층)

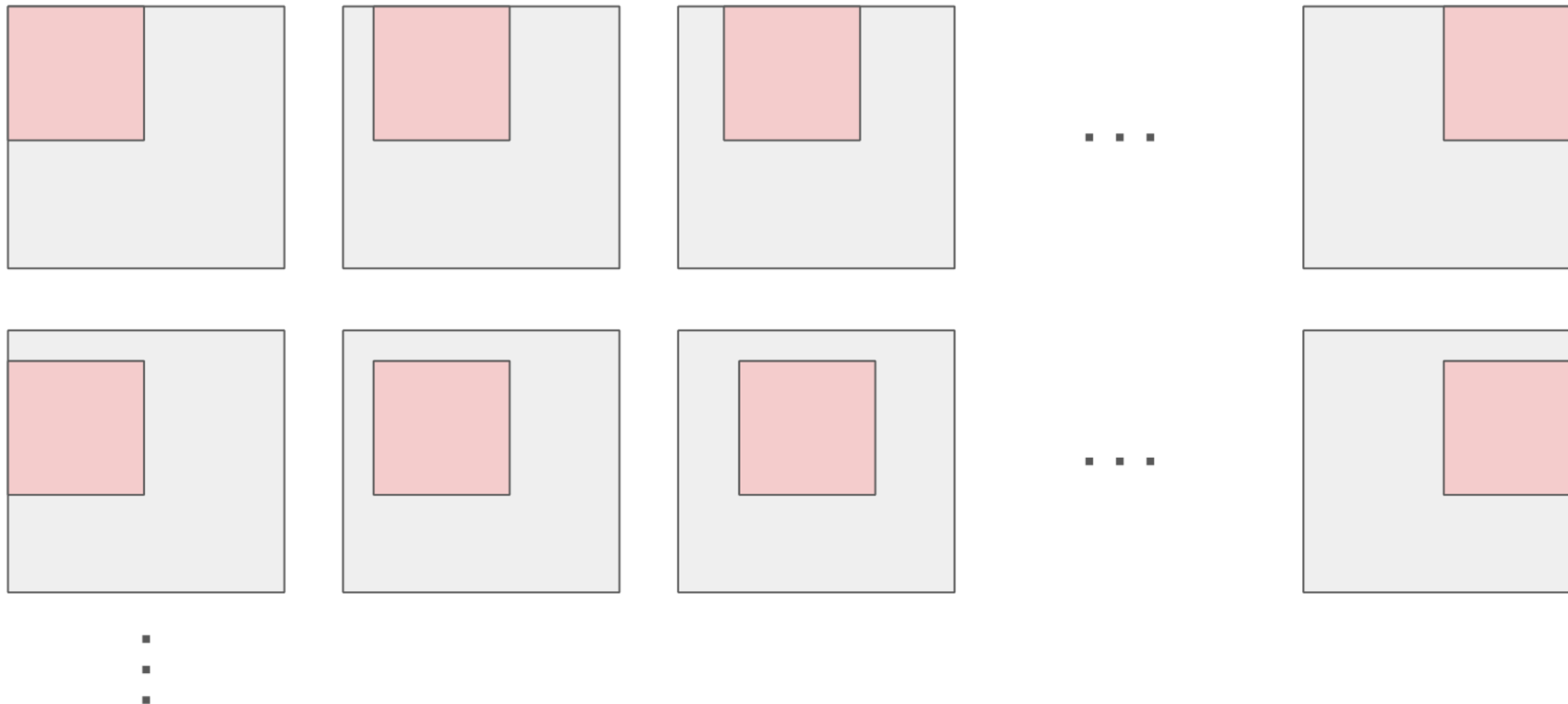
- 스트라이드





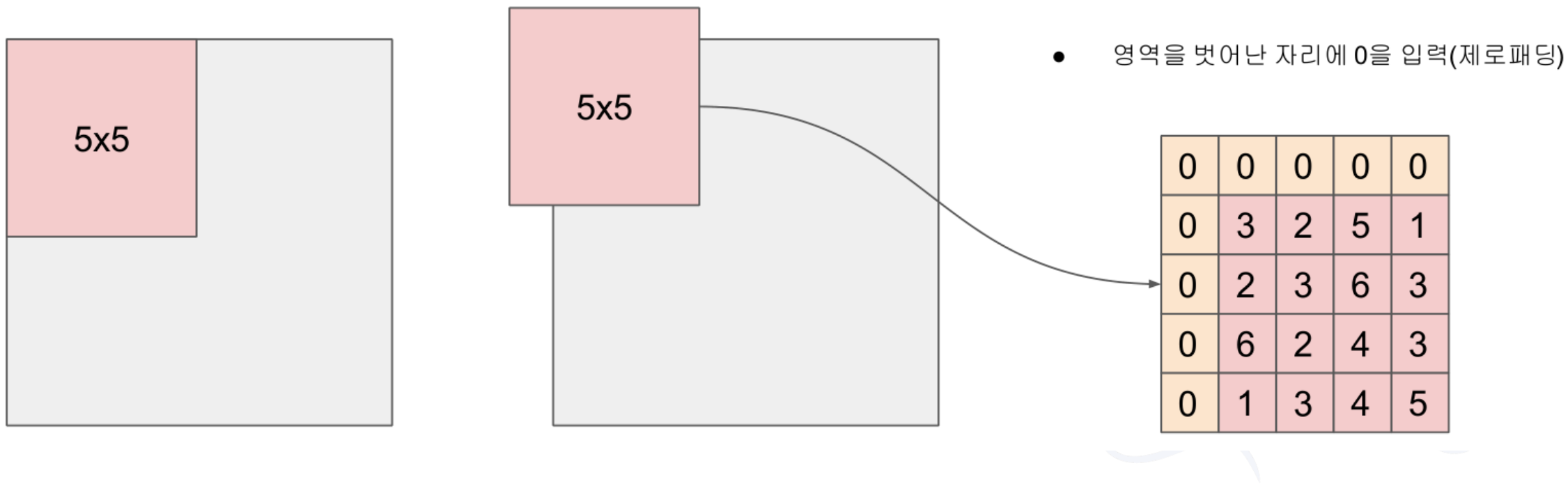
# Convolution Layer (합성곱 층)

- 스트라이드에 따른 이동



- 패딩

- 더 좋은 결과를 내기 위하여 윈도우를 이미지 영역 밖으로 확장할 때 사용



- Convolution Layer 에서의 연산

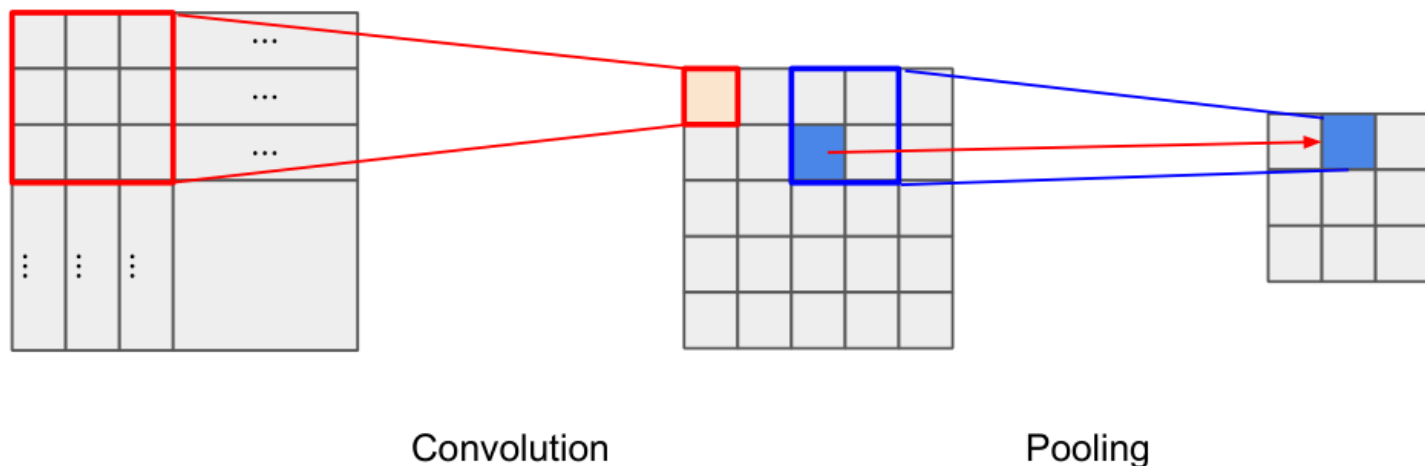
- Convolution, 즉 합성곱이란 이미지와 필터 사이에 정의되는 합성 곱 연산을 말함
- 이미지의 합성곱은 필터의 명암 패턴과 유사한 명암 패턴이 입력된 이미지의 어디에 있는지 검출하는 작용, 즉 필터가 나타내는 특징적인 명암 구조를 이미지로부터 추출하는 작용을 함
- 필터: 커널이라고도 함
  - 입력층의 윈도우를 은닉층의 뉴런 하나로 압축할 때, Convolution Layer에서는 윈도우의 크기만큼의 가중치와 1개의 편향 값(bias)을 적용
  - 예를 들어 윈도우의 크기가 5x5라면 5x5개의 가중치와 1개의 편향 값이 필요함
  - 이 5x5개의 가중치와 1개의 편향 값을 커널, 또는 필터라고 부름
  - 필터는 해당 은닉층을 만들기 위한 모든 윈도우에 공통으로 적용됨

- 필터 사용의 장점

- 가중치의 수를 줄임으로써 전체 연산량을 대폭 감소시킬 수 있음
- 예를 들어, 입력층의 크기가  $28 \times 28$  일때 기본 신경망의 경우  $28 \times 28 = 784$  개의 가중치를 찾아야 하지만 Convolution Layer 에서는  $5 \times 5$  개인 25 개의 가중치만 찾으면 됨
- 연산량의 대폭 감소로 학습이 더 빠르고 효율적으로 진행됨
- 단점
  - 복잡한 특징을 가진 이미지의 분석이 어려움
  - 보완책으로서 여러 개의 필터를 사용하며, 분석하고자 하는 내용에 따라 필터의 개수를 어떻게 정하는가 하는 것이 중요함

## • Convolution Layer 와 Pooling Layer 의 관계 및 역할

- 이미지 데이터, 즉 2 차원의 평면 행렬에서 지정한 영역의 값들을 하나의 값으로 압축
- 압축할 때
  - Convolution Layer: 가중치와 편향을 적용
  - Pooling Layer: 값들 중 하나를 선택해서 가져오는 역할



- Convolution Layer에서 해당 영역 내의 영상 위치가 다소 변경되더라도 인식할 수 있도록 함
- Max Pooling, Average Pooling이 많이 사용됨

- 국소 콘트라스트(대비) 정규화

- 자연물 이미지 등 주변의 조명, 카메라의 노출 등 환경 변화에 따라 이미지 전체의 밝기, 대비가 크게 변하는 경우 사용함

- 이미지 밝기 정규화의 방법

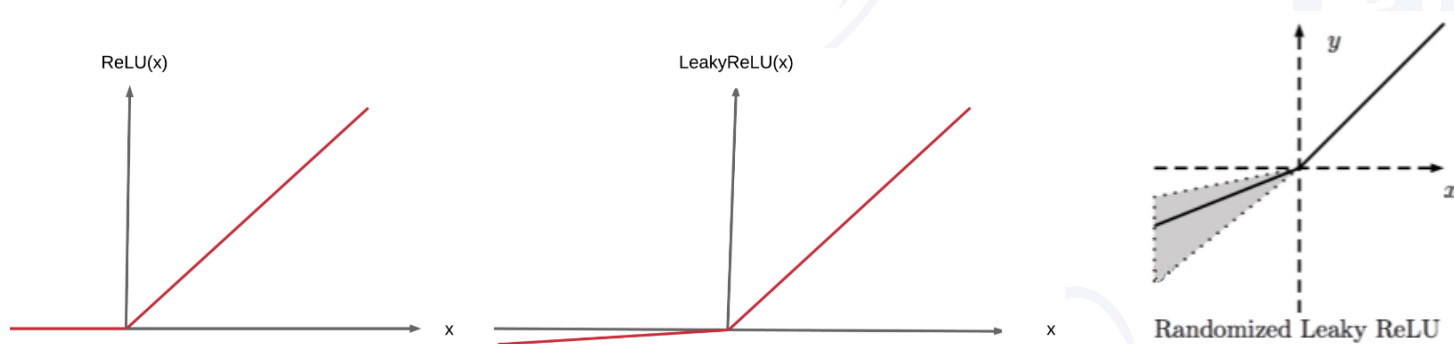
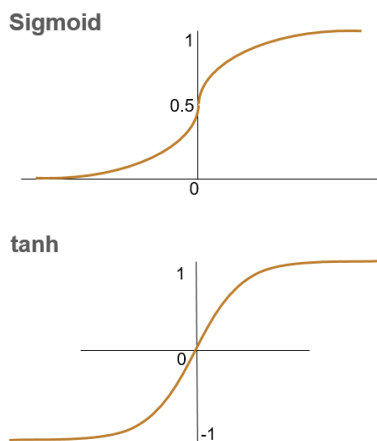
- 이미지의 집합(훈련 데이터)에 대한 통계치를 이용하여 이미지의 명암을 전체적으로 조절
  - LCN
    - 이미지 한 장, 한 장에 대하여 개별적으로 조절
    - 고정된 가중치를 사용하므로 학습 가능한 파라미터는 없음

- **활용도**

- 일반적으로 기존의 신경망에서 각 층별 연결에 사용되는 방식. 전결합층
- 모든 노드를 연결하므로 수많은 연산이 일어남
- CNN의 특징은 모든 노드를 결합하지 않음으로써 연산량을 줄여 효율성을 높이는 방식
- 그럼 왜 사용하는가?
  - 모든 노드를 연결하므로 1차원배열로 표시됨 → 이미지의 공간정보가 사라짐
  - 최종 결과값은 분류 결과 도출      결국 마지막에 도출된 분류결과 Label을 선택하여야 함
  - 최종 결과를 분류하기 위한 기반 정보는 모두 가지고 있어야 분류를 위한 SoftMax 함수를 사용할 수 있음
- 필수는 아니며 Convolution Layer의 결과를 그대로 사용할 수도 있음

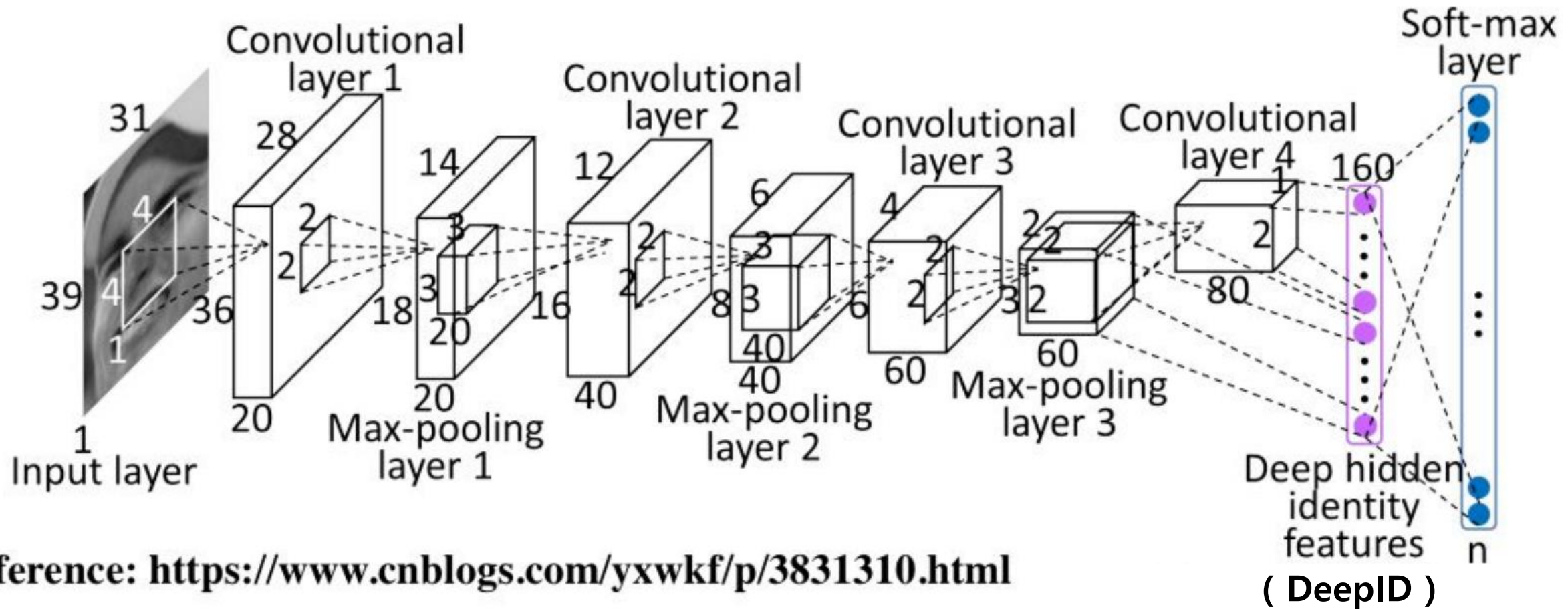
## • 활성화 함수

- 실제 신경망에서는 우리 몸에서 반응할 필요가 있는 수준 까지만 신호를 전달하고 나머지의 신호는 무시 → 비선형적 특징
- 합성곱 연산은 입력과 가중치로 이루어진 연산 → 선형성을 가짐 → 비선형 특성 부여를 위하여 활성화 함수가 필요함
- 최근 많이 사용되는 활성화 함수는 ReLU(Rectified Linear Unit) 계열의 함수

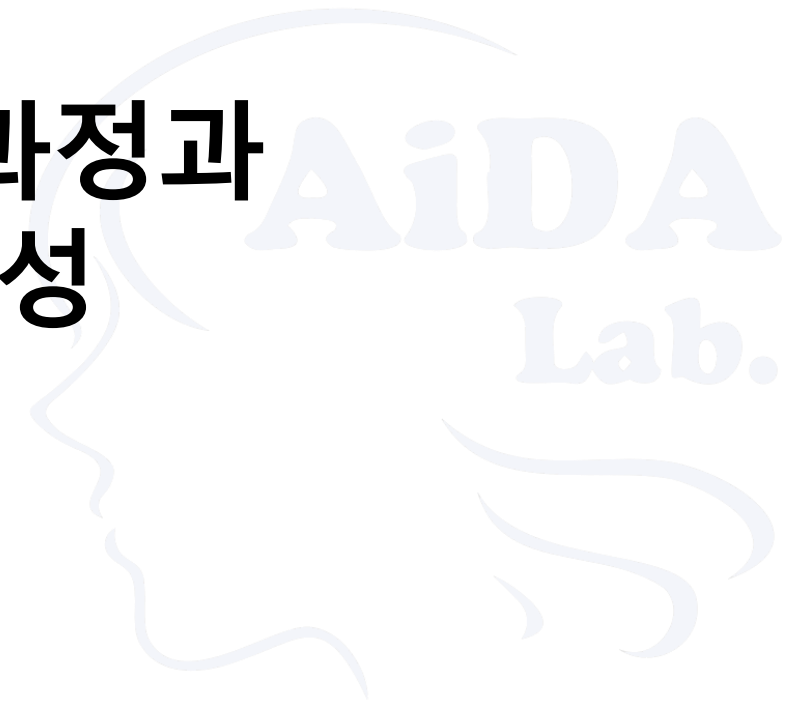




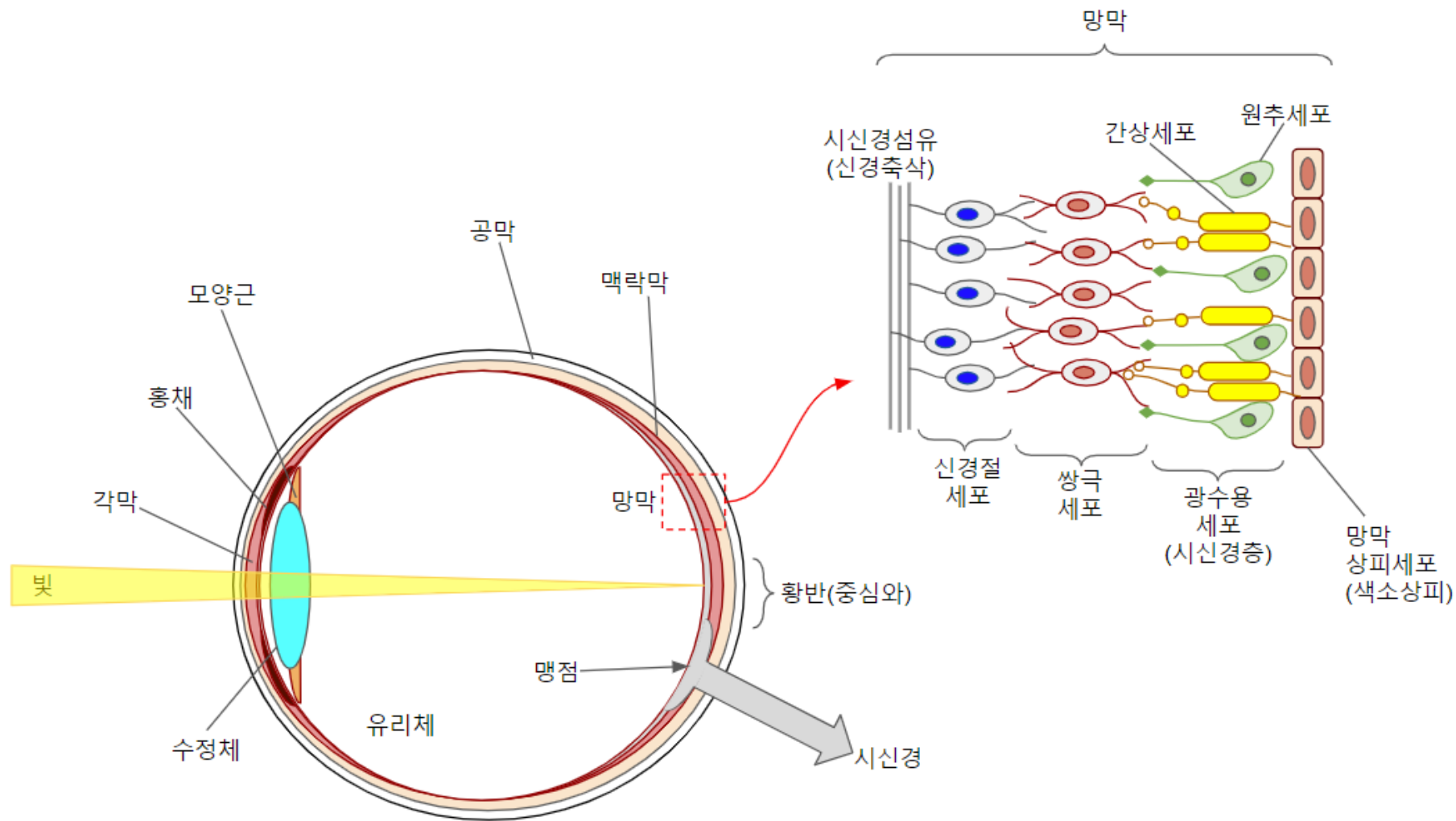
# CNN 전체 구조의 예시



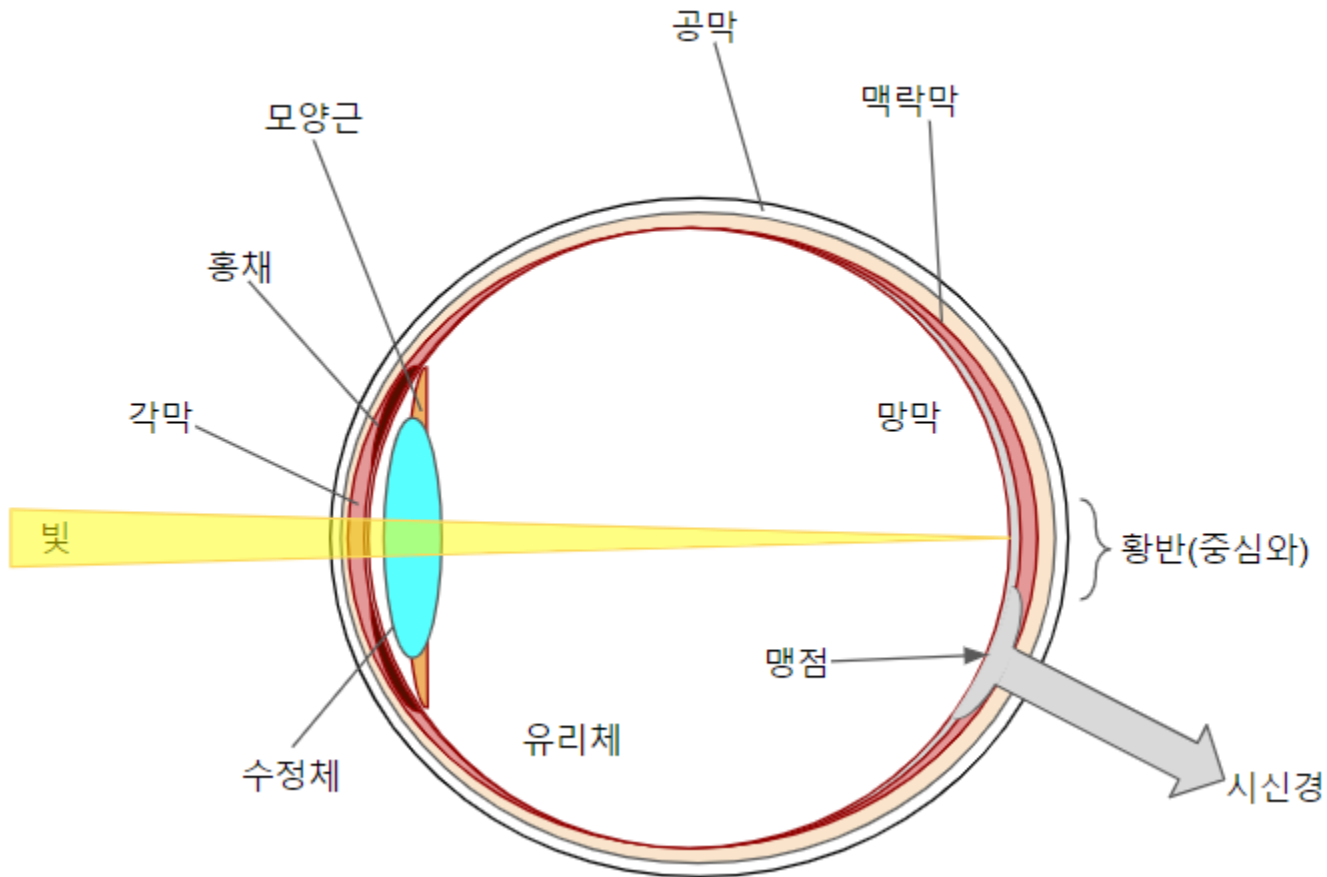
# 뇌의 시각정보 처리 과정과 CNN 모델의 연관성



# 뇌의 시각정보 처리 구조: 눈과 망막의 구조

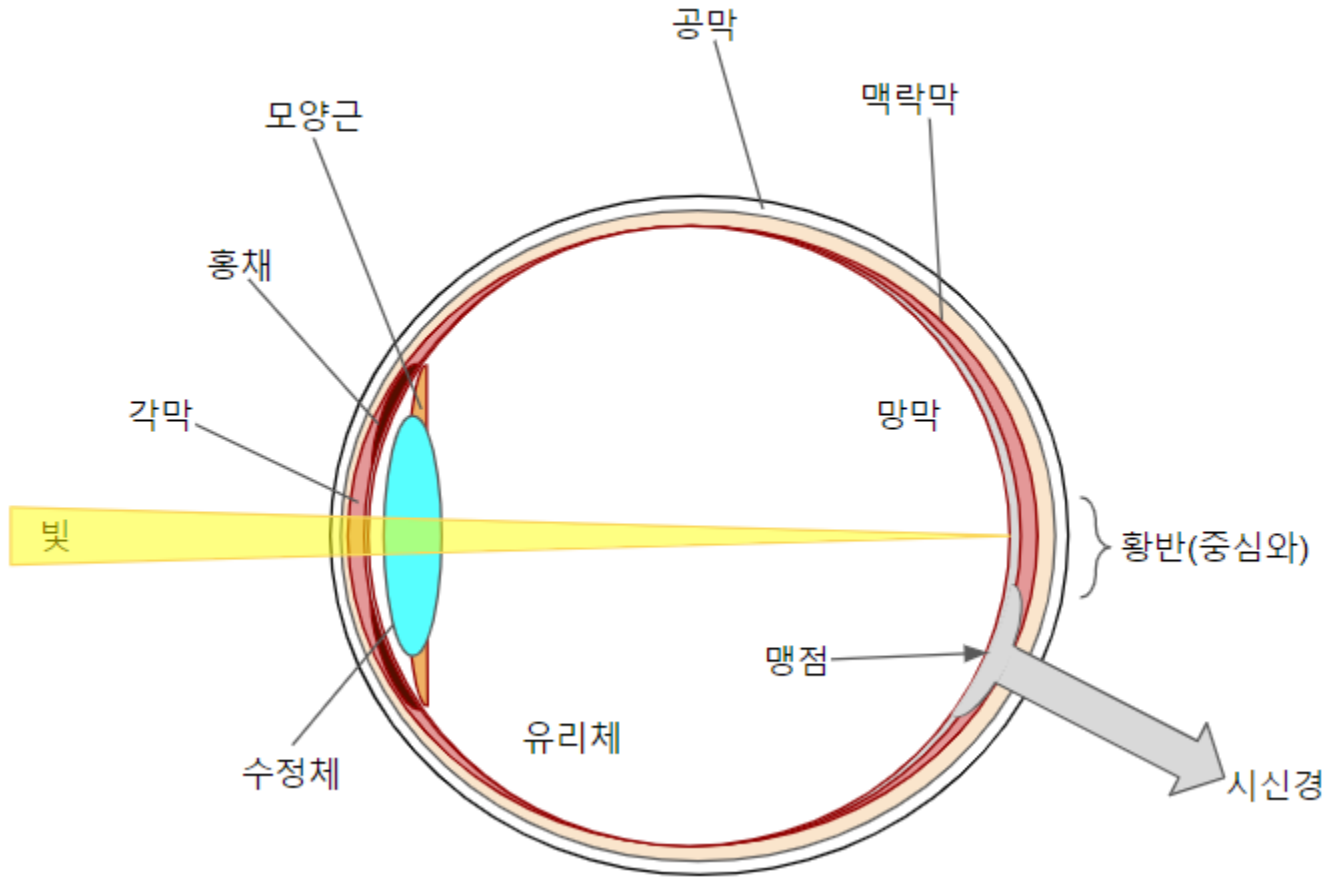


# 뇌의 시각정보 처리 구조: 눈과 망막의 구조



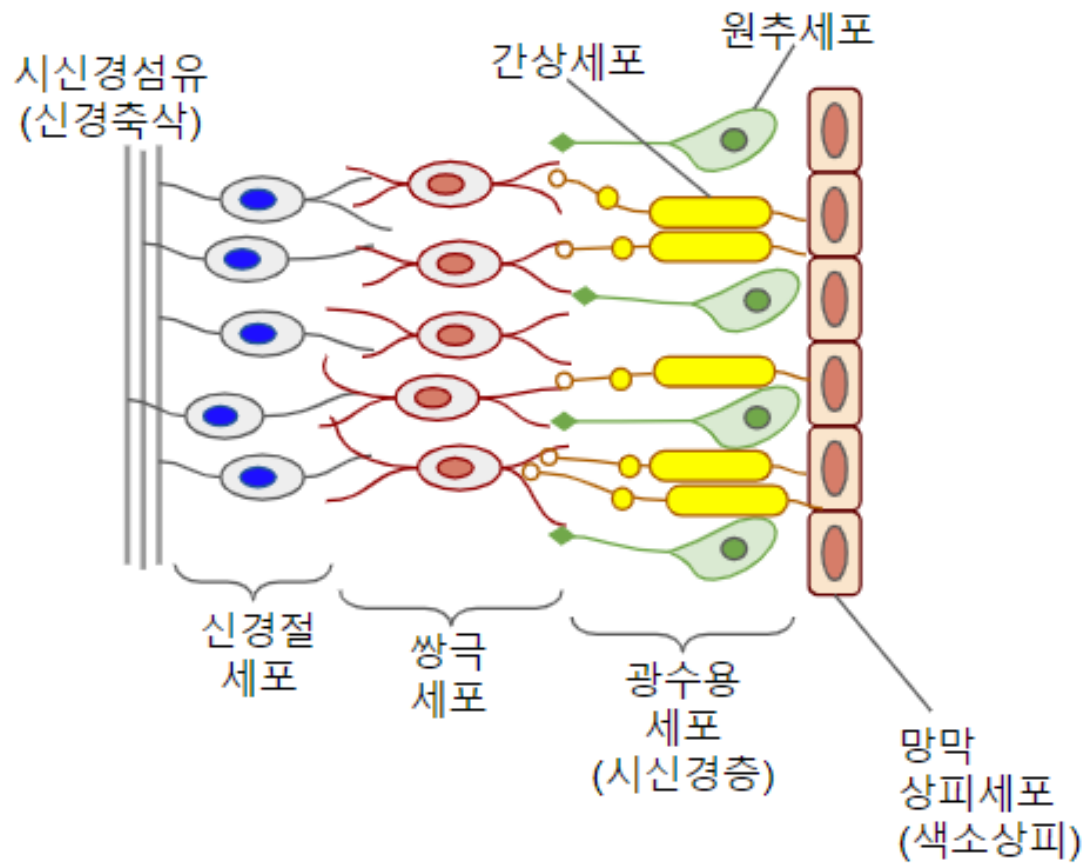
- 아주 좁은 영역이지만 가장 세밀하고 선명하게 빛을 인식함
- 중심와에는 원추세포가 가장 밀집되어 있음
- 중심와에서 멀어질수록 원추세포의 밀집도는 떨어지고 간상세포가 늘어남

# 뇌의 시각정보 처리 구조: 눈과 망막의 구조



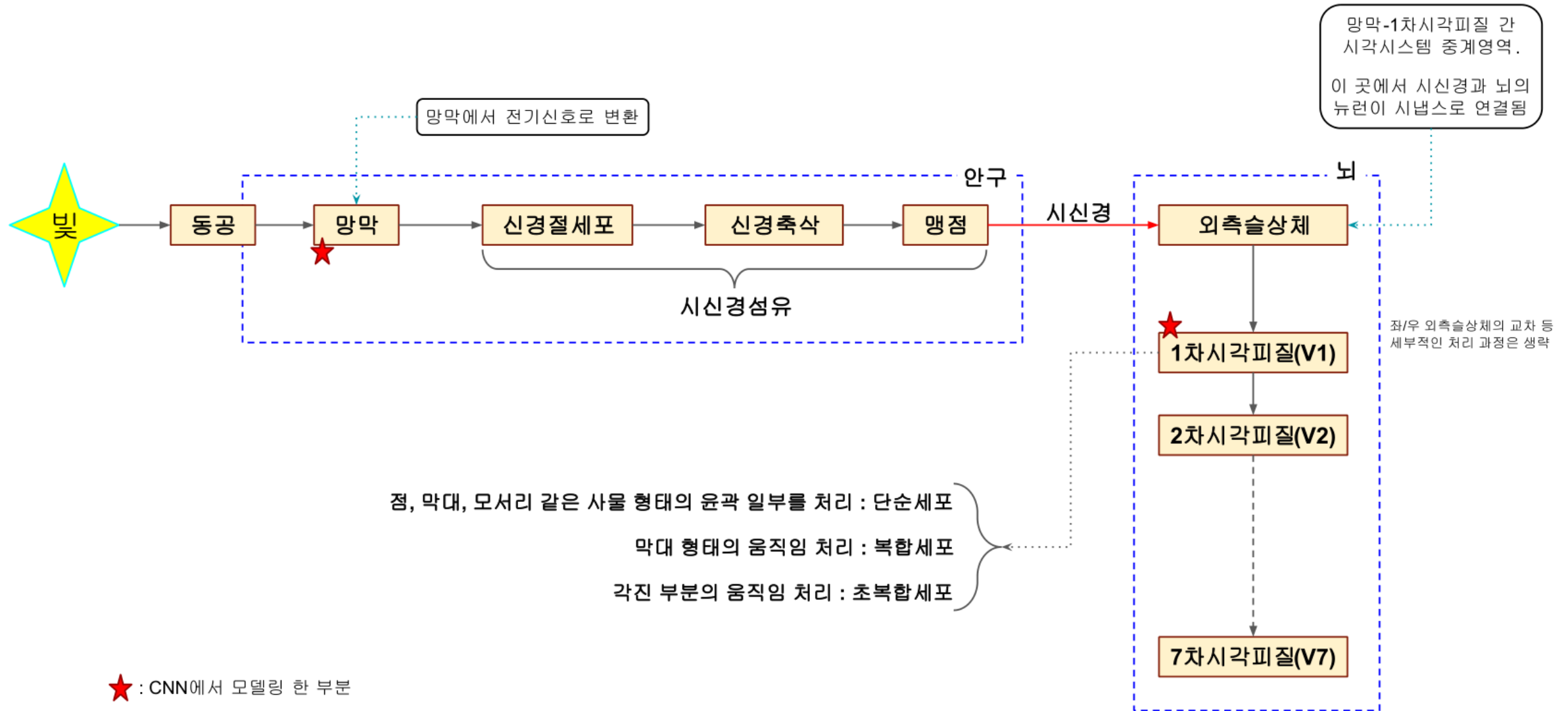
- 눈의 잘못된 진화(?)와 노이즈 캔슬링 기능  
→ 인간의 몸에서 찾아볼 수 있는 전자/기계적 예를 볼 수 있다.

# 뇌의 시각정보 처리 구조: 눈과 망막의 구조



- 광수용세포는 빛에 대해 화학적으로 반응하는 광색소 함유
- 색채를 지각하는 원추세포, 명암을 탐지하는 간상세포
- 각 원추세포마다 다른 광색소를 함유함 (적색, 녹색, 청색)
- 빛이 광수용세포에 닿으면 광색소의 생화학적 특성에 따라 전위 변화가 발생하고 이온 투과성을 변경시킴
- 이온 투과성의 변화에 따라 광수용세포에서 신경전달물질이 분비되며 이로 인하여 전기 신호가 발생, 전달됨

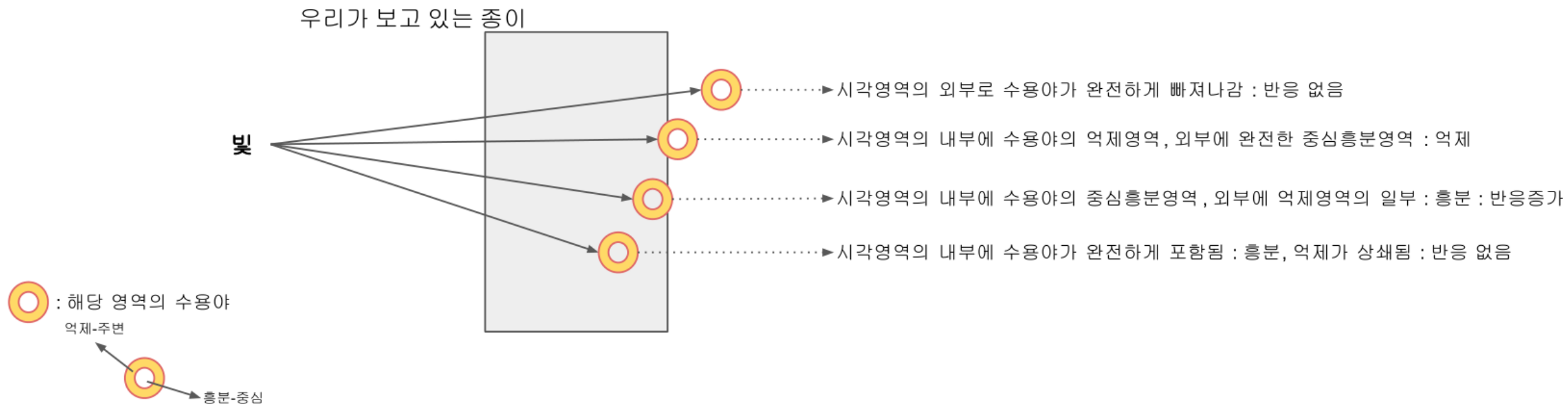
# 뇌의 시각정보 처리 구조: 시각정보의 흐름



- 안구에서 신경절 세포는 마지막 출력에 해당하는 영역, 뉴런과 동일한 방식으로 동작함
- 망막에서 전기신호로 변환된 시각 정보는 신경절 세포의 반응률에 영향을 미침
- 신경절 세포의 반응률에 영향을 미치는 망막 표면의 영역을 해당 세포의 수용야(Receptive Field) 라고 하며 “ 중심흥분 + 주변억제 ” 와 “ 중심억제 + 주변흥분 ” 의 두 가지 형태가 존재함



# 뇌의 시각정보 처리 구조: 시각 피질



- **시각피질에서의 시각정보 처리**

- **1 차 시각피질의 세포들**

- 단순세포: 국소적인 영역을 보고, 단순한 패턴에 자극을 받는 세포  
→ 점, 막대, 모서리 같은 사물 형태의 윤곽 일부를 처리
- 복합세포: 넓은 영역을 보고, 복잡한 패턴에 자극을 받는 세포 → 막대 형태의 움직임 처리
- 초복합세포: 각진 부분(모서리)의 움직임 처리

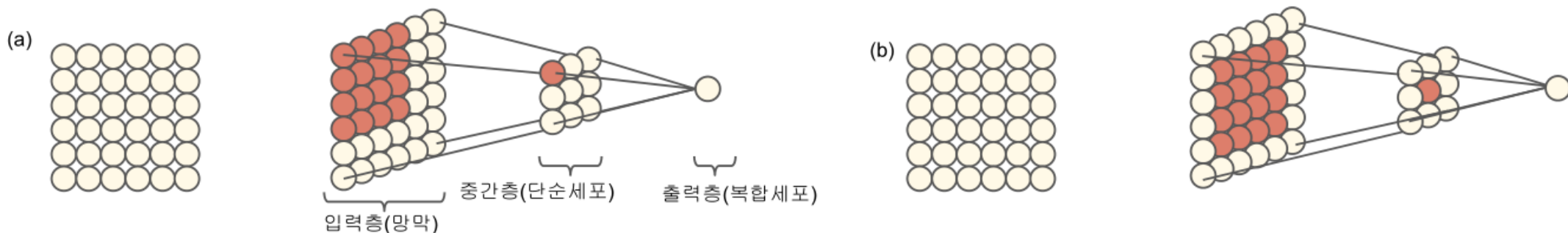
- **각 세포들은 수직으로 세워놓은 기둥 형태로 조직되어 있음( 시각피질에서만 이런 형태)**

- **각 세포들은 서로 다른 방향의 자극에 대해 반응함**

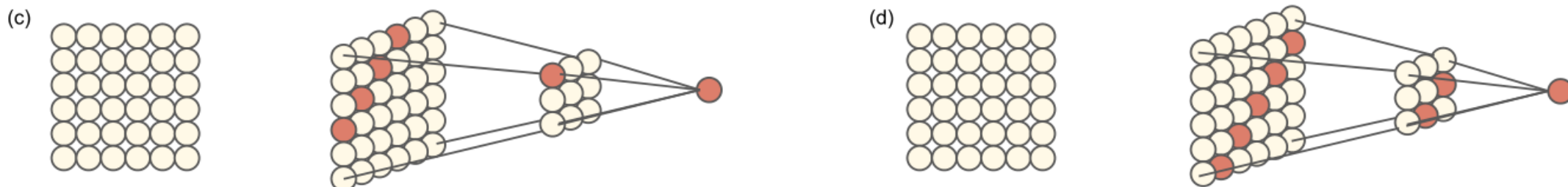
- 시각피질에서 이러한 기둥 조직에 의해 방향성 정보를 부호화 하는 것이 시각적 공간을 뉴런에 의해서 재구성하는데 매우 중요한 역할을 하는 것으로 추정됨
- 세포기둥은 세포들의 단순한 집합이 아니라 역동적 기능 단위이다

# 뇌의 시각정보 처리 구조: 시각 피질

- 단순세포: 엄격한 위치 선택성을 가짐 (정확한 입력패턴에 반응)
- 복합세포: 입력패턴을 조금 벗어나도 반응함
- 단순세포와 복합세포의 반응성을 모형화 하면:



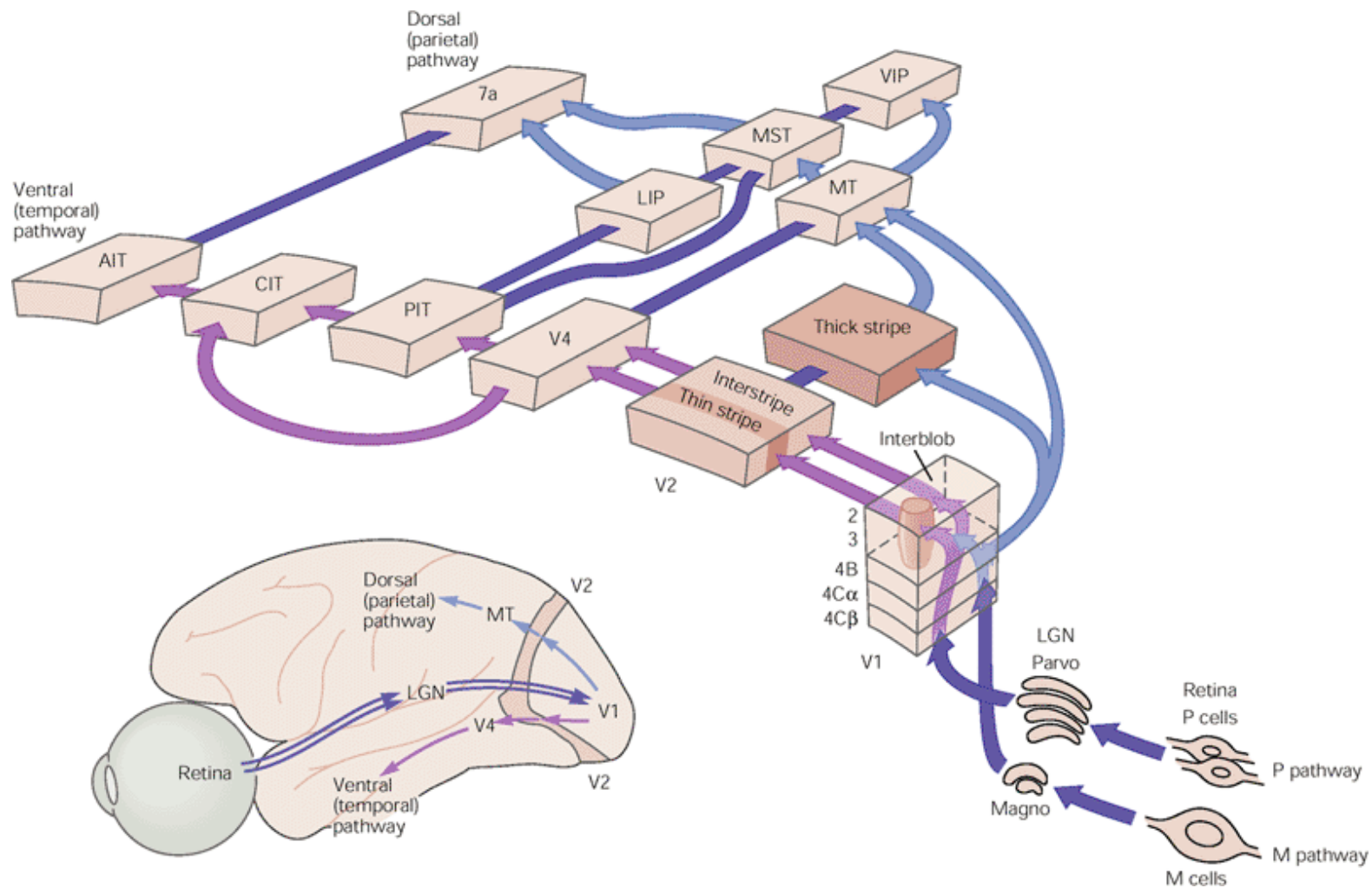
중간층 각 유닛은 입력층의  $4 \times 4 = 16$ 개의 유닛하고만 결합, 출력층 유닛은 중간층 유닛  $3 \times 3 = 9$ 개의 유닛 중 하나라도 활성화되면 활성화



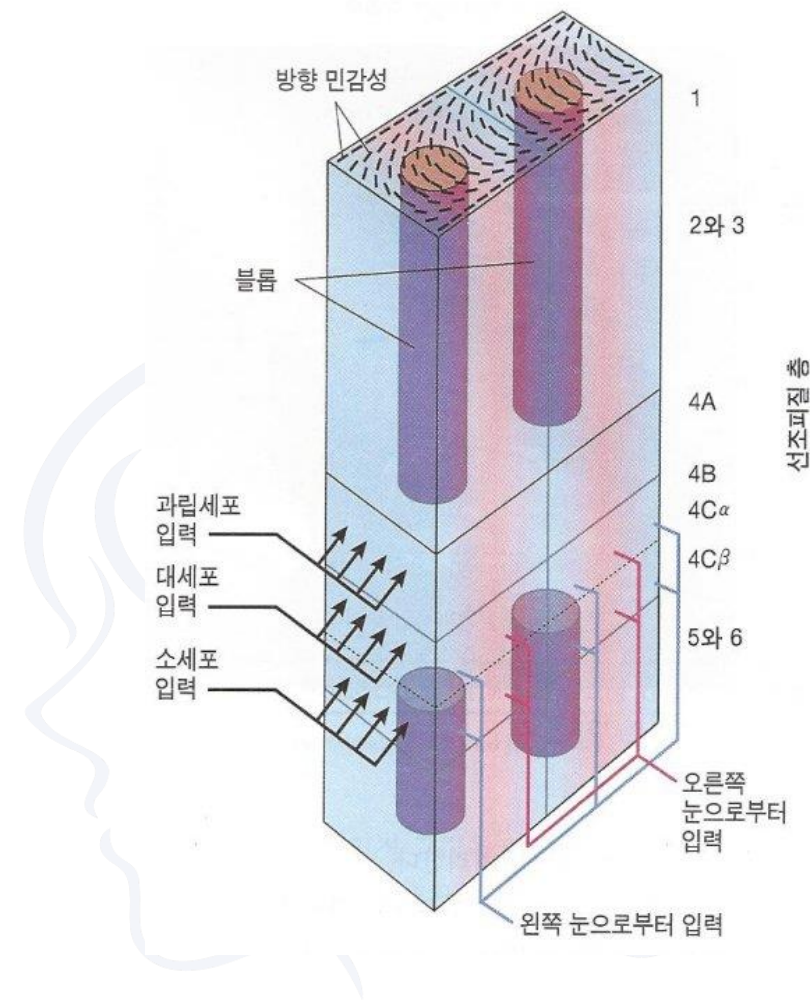
중간층의 유닛은 입력 패턴의 위치에 민감하지만, 출력층의 유닛은 범위 안에서 둔감함(하나라도 활성화되면 자기도 활성화)

- 눈과 시각피질 사이의 시각정보 처리 방식에서 개념 도입
  - 망막에서의 빛 인식 영역 → 입력층
  - 시각피질의 단순세포에 의한 입력패턴 대응 영역 → 중간층
  - 시각피질의 복합세포에 의한 활성화 및 차원축소 영역 → 출력층
  - 해당 영역에 대한 수용야(감수영역이라고도 함) → 필터
  - 수용야의 영역 크기 → 윈도우
  - 수용야의 이동 범위 → 스트라이드
  - 수용야의 영역 처리 → 패딩
- 시각 처리 신경망은 단순세포와 복잡세포가 층을 이루어 구성되었고, 층간 연산에 따라 동작하는 것으로 관찰되고 있으며, CNN 모델은 이를 반영하고 있음

# 시각의 전달 경로 (CNN과 비교해보자)



일차시각피질의 모듈 중 하나



2차 출처: <http://www.seehint.com/word.asp?no=13222>