

# CNN

## CNN 네트워크의 동작 구조와 NezNet 이해하기

“이미지는 단순한 픽셀의 집합이 아니다.”

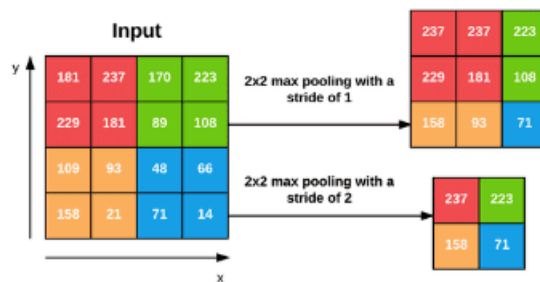
인공지능이 이미지를 이해하도록 만드는 핵심 기술, 바로 **CNN(Convolutional Neural Network)** 입니다.

이번 글에서는 CNN의 동작 원리를 그림과 함께 알아보고, 이를 기반으로 **NezNet** 같은 네트워크를 어떻게 설계할 수 있는지 살펴보겠습니다.

### 1. CNN이란 무엇인가?

CNN(합성곱 신경망)은 이미지나 영상처럼 **공간적 구조(spatial structure)** 를 가진 데이터를 처리하기 위해 만들어진 신경망입니다.

#### CONVOLUTIONAL NEURAL NETWORKS (CNNs) AND LAYER TYPES



일반적인 신경망이 모든 픽셀을 독립적으로 다룬다면,

CNN은 **인접한 픽셀 간의 관계**를 이용해 '엣지', '모양', '패턴'을 학습합니다.

즉, **시각적 특징(feature)** 을 점진적으로 추출하는 구조라고 볼 수 있습니다.

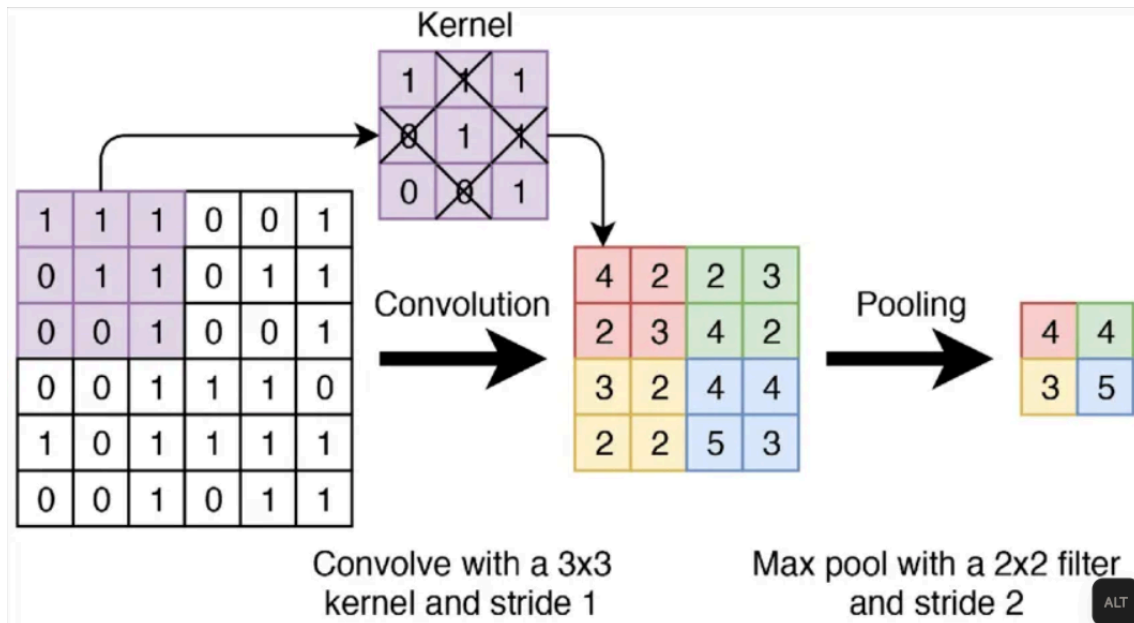
### 2. CNN의 기본 구성

CNN은 보통 다음과 같은 흐름으로 구성됩니다.

단계	설명
입력층(Input Layer)	이미지 입력 (예: 224×224×3)
합성곱층(Conv Layer)	지역 특징 추출, 필터(커널) 적용
활성화 함수(Activation)	비선형성 추가 (ReLU 등)
풀링층(Pooling)	해상도 축소, 연산량 감소
정규화/드롭아웃	학습 안정화, 과적합 방지
완전연결층(FC Layer)	특징을 종합해 최종 예측
출력층(Output Layer)	Softmax 등으로 결과 확률화

### 3. 핵심 단계별 동작 원리

#### (1) 합성곱층 – 특징 추출의 핵심

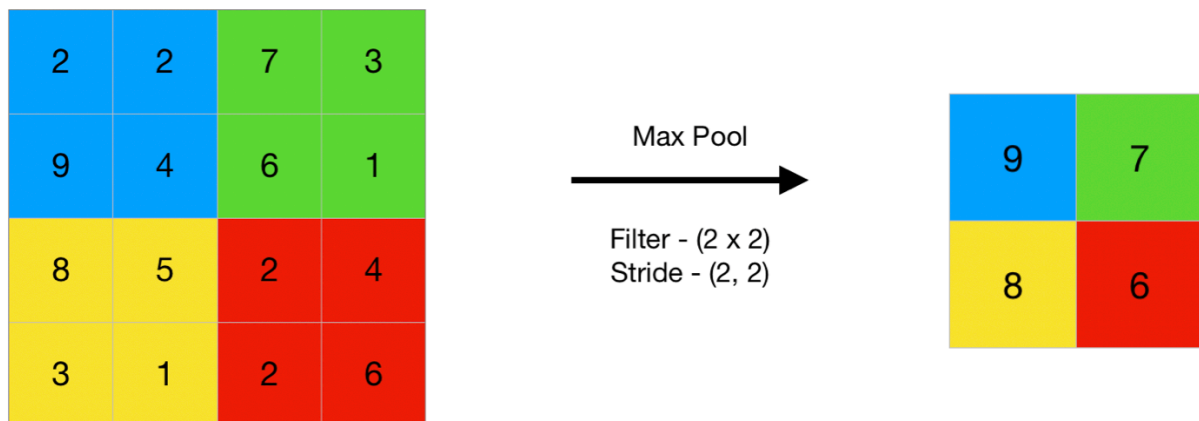


- 입력 이미지에 **필터(kernel)**를 슬라이딩하며 곱셈·합산 연산을 수행합니다.
- 각 필터는 특정한 패턴(예: 엣지, 텍스처 등)에 반응하게 학습됩니다.
- 결과적으로 **특징맵(feature map)**이 생성됩니다.

#### (2) 활성화 함수 – 비선형 변환

- ReLU(Rectified Linear Unit):  $f(x) = \max(0, x)$   
 $f(x) = \max(0, x)$
- 이 과정을 통해 모델이 단순한 선형 결합을 넘어, 복잡한 패턴까지 학습할 수 있습니다.

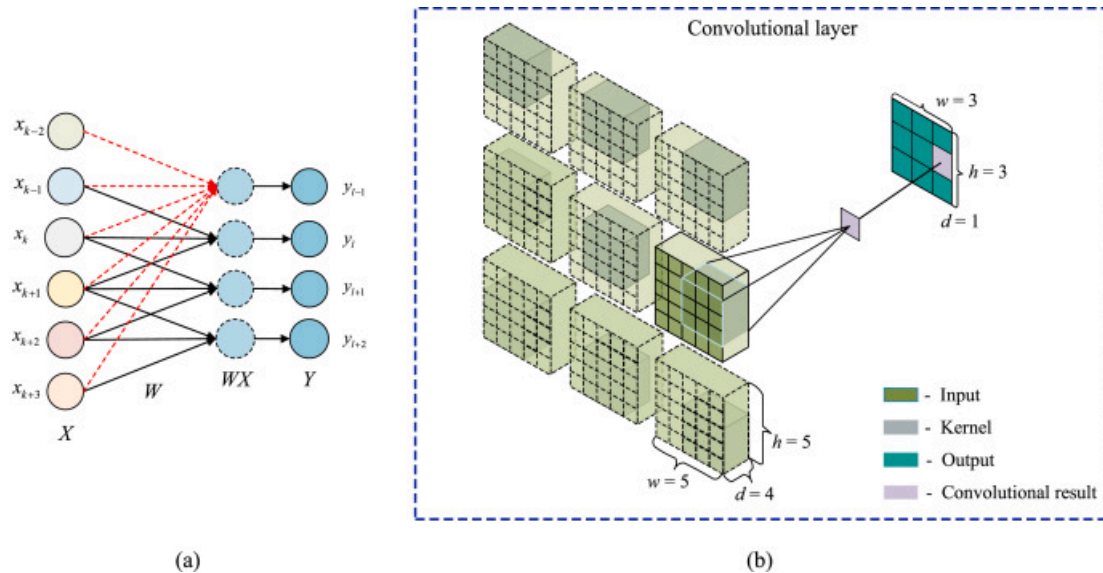
#### (3) 풀링층 – 차원 축소와 위치 불변성



- **Max Pooling:** 가장 큰 값을 선택

- **Average Pooling:** 평균값을 선택
- 특징의 핵심 정보를 유지하면서 계산량을 줄이고, 물체의 위치 변화에 강인한 모델을 만듭니다.

#### (4) 완전연결층 – 추출된 특징을 종합



- 앞 단계에서 추출한 모든 특징을 펼쳐서(flatten) **분류 또는 회귀 예측**을 수행합니다.
- Softmax를 통해 각 클래스의 확률을 출력하는 구조가 일반적입니다.

## 4. 학습의 흐름 – Forward & Backward

CNN의 학습은 다음 과정을 반복합니다.

1. **순전파(Forward):** 입력 → 출력
2. **손실 계산(Loss)**
3. **역전파(Backpropagation):** 오차를 이용해 가중치 조정
4. **업데이트(Update):** 최적화 알고리즘(SGD, Adam 등)으로 학습

## 5. 계층적 특징 학습 (Hierarchical Feature Learning)

CNN은 깊이가 깊어질수록 더 복잡한 특징을 학습합니다.

층 깊이	학습되는 특징
초반부	엣지, 코너, 간단한 패턴
중간부	텍스처, 모양
후반부	구체적 객체(눈, 입, 자동차 등)

## 6. NezNet에 적용하기

현재 NezNet의 구체적인 구조는 공개되어 있지 않지만,

CNN의 기본 원리를 기반으로 NezNet을 설계하거나 분석할 때 아래 요소를 고려할 수 있습니다.

항목	질문
입력 사양	입력 크기, 채널 수는?
목표(Task)	분류, 검출, 세분화 중 어느 용도인가?
Conv 블록 구성	각 층의 필터 수, 크기, 활성화 함수는?
Downsampling 방식	MaxPooling, Strided Conv 중 어느 것 사용?
정규화 및 잔차 연결	BatchNorm, Residual Block, Skip Connection 존재 여부
출력 구조	FC, GAP(Global Avg Pooling), Softmax 등
최적화 전략	손실 함수, 옵티마이저, 데이터 증강 방식 등

## 7. NezNet 예시 구조 (가정 예시)

입력:  $224 \times 224 \times 3$  이미지  
 ↓  
 Conv( $7 \times 7$ , stride=2) + ReLU → MaxPool( $3 \times 3$ )  
 ↓  
 Conv Block  $\times 4$  (채널 수 증가)  
 ↓  
 Global Average Pooling  
 ↓  
 Fully Connected → Softmax  
 ↓  
 출력: 클래스 확률

만약 NezNet이 ResNet이나 Inception처럼 잔차 연결(Residual Connection) 또는 병렬 필터 구조를 포함한  
 다면,  
 단순 CNN보다 효율적이고 깊은 학습이 가능할 것입니다.

## 결론

CNN은 단순히 이미지를 인식하는 것이 아니라,

**공간적 패턴을 단계적으로 학습하는 신경망 구조**입니다.

NezNet 역시 이런 CNN의 기본 원리를 바탕으로 설계된 네트워크로 보입니다.

앞으로 NezNet의 구체적인 구조가 공개된다면,

그 내부의 **합성곱 블록 구조, 채널 구성, 활성화 방식**을 분석해

어떤 비전을 중심으로 설계되었는지 깊이 이해할 수 있을 것입니다.