



**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**  
Posts and Telecommunications Institute of Technology

# Nhận dạng chữ viết và hình dạng đơn giản bằng mạng neural

Nhóm thực hiện: 08

Thời gian: 17/11/2025

# Mục lục

---

- 1** Tổng quan đề tài
- 2** Cơ sở lý thuyết
- 3** Ứng dụng
- 4** Mô phỏng

>>> Nội dung tiếp theo

## Tổng quan đề tài

---

- Giới thiệu
- Mục tiêu
- Phạm vi

# Tổng quan đề tài

## Giới thiệu

---

---

- **Nhận dạng chữ viết và hình dạng** là bài toán quan trọng trong **trí tuệ nhân tạo và thị giác máy tính**.
- Ứng dụng thực tế :
  - Tự động đọc tài liệu, hóa đơn, chứng từ.
  - Hỗ trợ nhập liệu nhanh, chính xác.

# Tổng quan đề tài

## Giới thiệu

---

---

- Một số mô hình học máy truyền thống :
  - Độ chính xác chưa cao
  - Phụ thuộc nhiều vào bước **trích xuất đặc trưng thủ công**
- Hướng tiếp cận của nhóm:
  - Ứng dụng Mạng neuron tích chập (CNN – Convolutional Neural Network)
  - Tự động trích xuất đặc trưng ảnh
  - Cho kết quả nhận dạng **chính xác và ổn định hơn.**

# Tổng quan đề tài

## Mục tiêu

---

---

- Nghiên cứu và ứng dụng **CNN** trong nhận dạng chữ viết & hình dạng.
- Hiểu rõ quá trình **xử lý ảnh đầu vào, trích chọn đặc trưng, và phân loại bằng mạng neural**.

# Tổng quan đề tài

Phạm vi

---

---

- Nhận dạng offline : xử lý ảnh ký tự hoặc hình dạng đã được viết/in sẵn
- Đối tượng nhận dạng :
  - Chữ viết tay
  - Hình dạng cơ bản (Hình tròn, hình chữ nhật, hình tam giác) trong ảnh.

>>> Nội dung tiếp theo

## Cơ sở lý thuyết

---

- Khái niệm CNN
- Các lớp của CNN
- Quy trình hoạt động của CNN

# Cơ sở lý thuyết

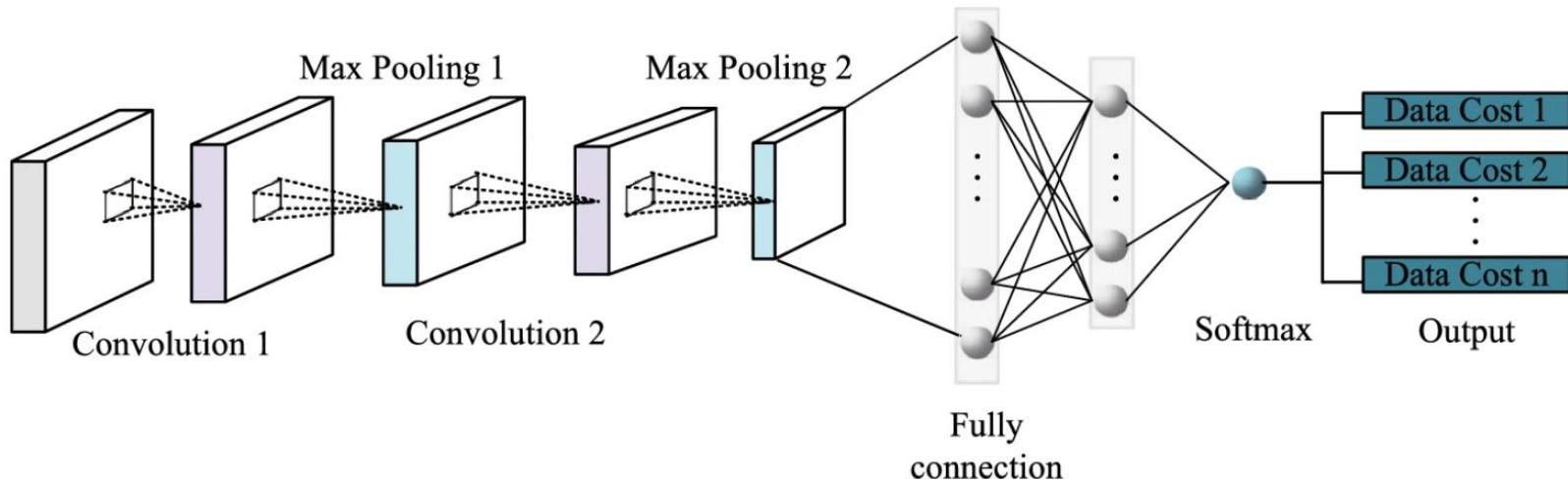
## Giới thiệu về CNN (Convolutional Neural Network)



- Là một trong những mô hình Deep Learning cực kỳ tiên tiến, bởi chúng cho phép xây dựng những hệ thống có độ chính xác cao và thông minh.
- CNN có rất nhiều ứng dụng, đặc biệt là những bài toán cần nhận dạng vật thể (object) trong ảnh.

# Cơ sở lý thuyết

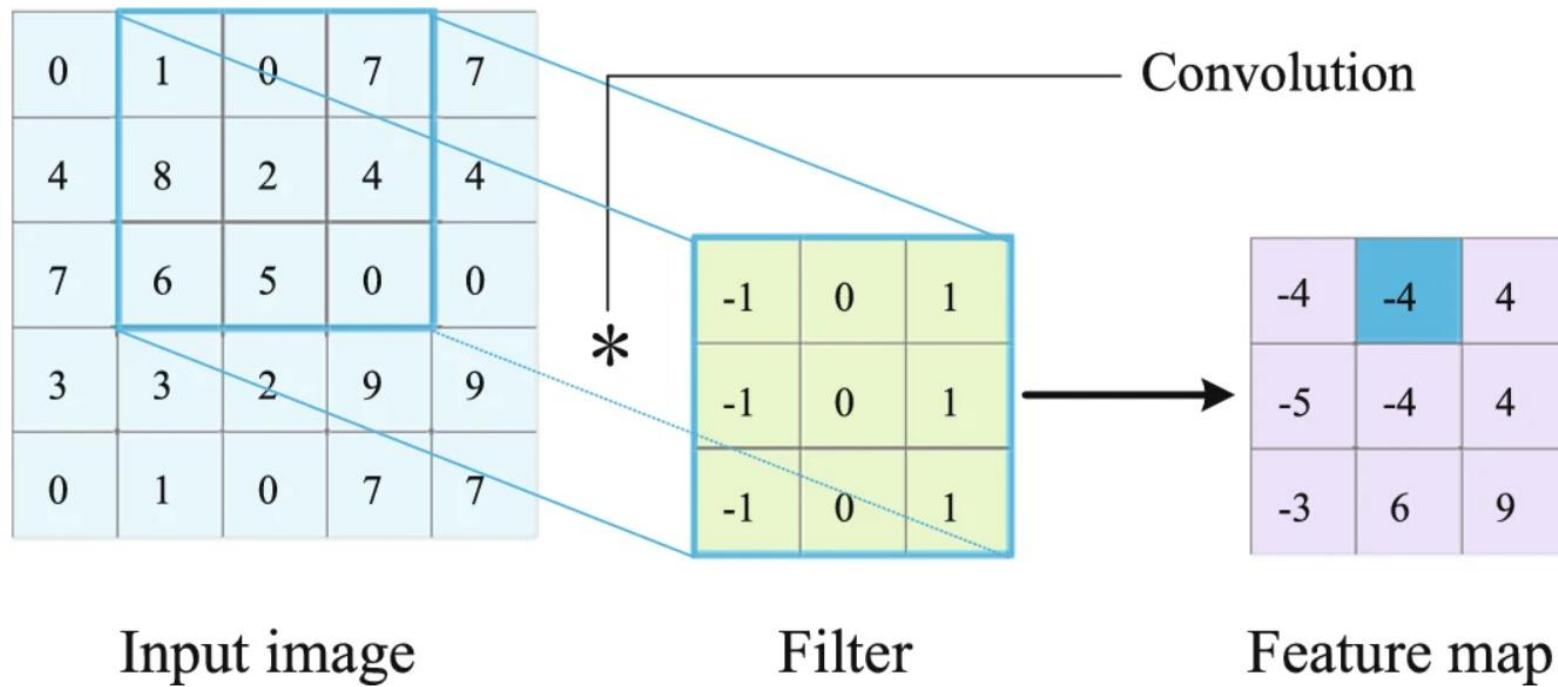
## Các thành phần cơ bản của CNN



- Về mặt tổng quan CNN bao gồm : convolution layer, activation layer, pooling layer và fully connected layer.
- Các lớp được kết nối đầy đủ được sử dụng kết hợp với các hàm kích hoạt để thực hiện các tác vụ như phân loại hoặc hồi quy trên các tính năng được trích xuất.

# Cơ sở lý thuyết

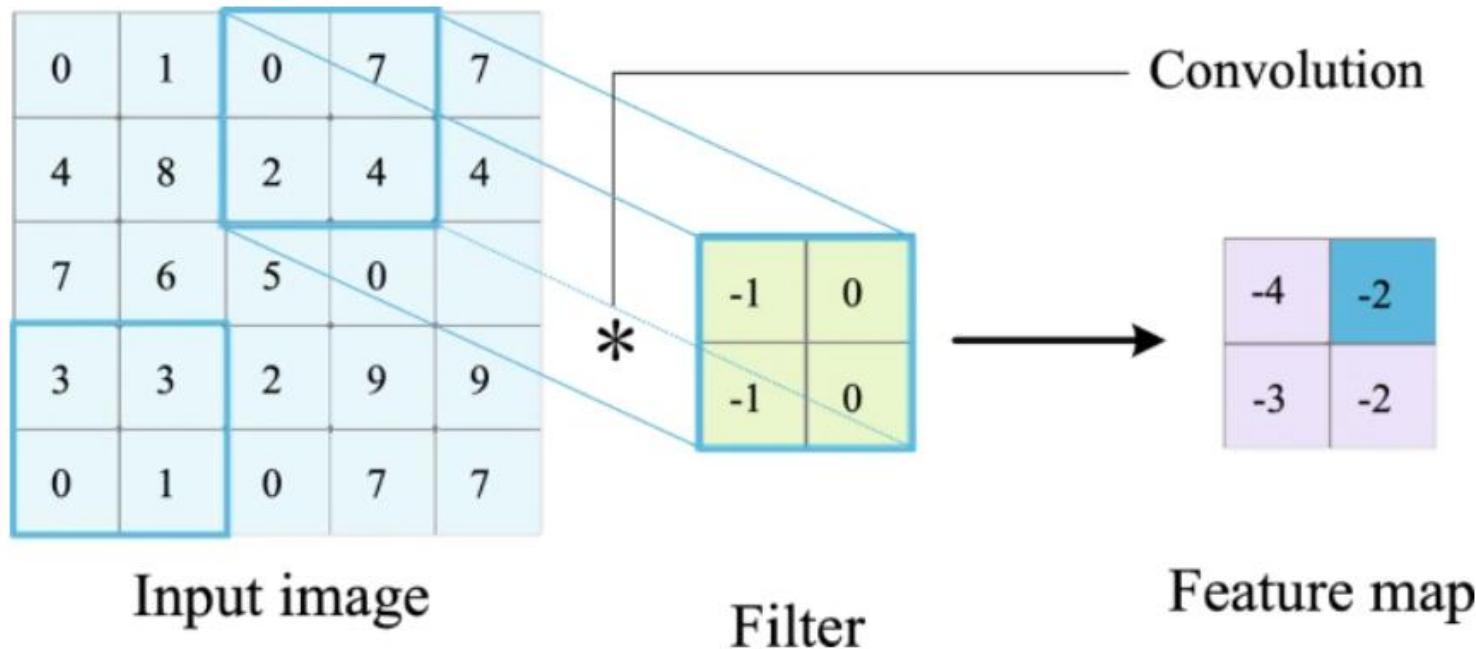
## Convolution layer



- Dùng để **trích xuất đặc trưng cục bộ** từ ảnh đầu vào.
- Sử dụng **kernel (bộ lọc)** kích thước nhỏ (thường  $3 \times 3$  hoặc  $5 \times 5$ ).
- Bộ lọc **trượt (stride)** qua ảnh và **tích chập** với vùng ảnh tương ứng.
- Giúp hỗ trợ phát hiện các đặc trưng như **cạnh, góc, hình dạng**.

# Cơ sở lý thuyết

## Convolution layer

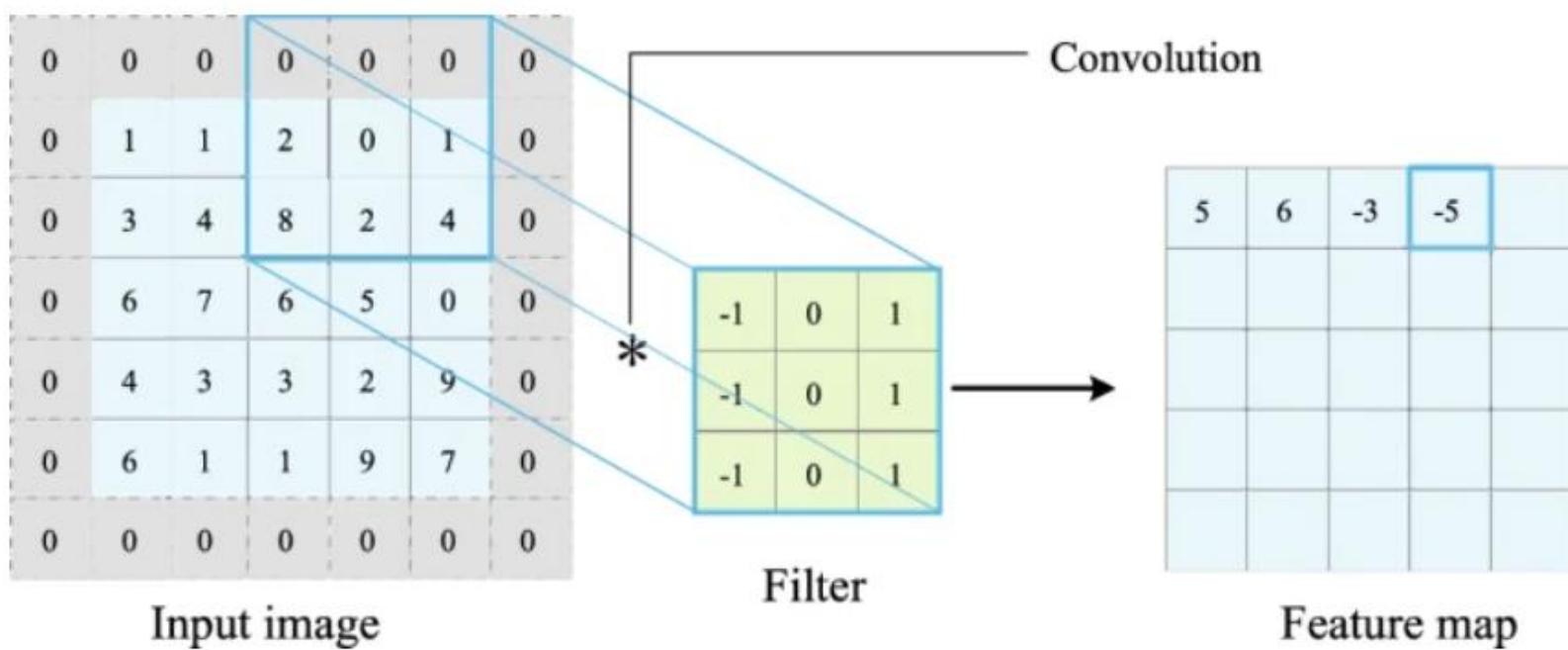


Convolution procedure (stride=(2,3))

- **Stride** (bước nhảy) là **số ô mà bộ lọc (kernel)** di chuyển **mỗi lần trượt** qua ảnh đầu vào.
- **Stride nhỏ (1)**: giữ nhiều thông tin chi tiết → ảnh đầu ra lớn hơn.
- **Stride lớn (2, 3)**: giảm kích thước nhanh → ít thông tin hơn, xử lý nhanh hơn.

# Cơ sở lý thuyết

## Convolution layer



Padding

- **Padding** (đệm) được sử dụng để ngăn chặn kích thước bản đồ đặc trưng bị thu hẹp sau mỗi lớp.

# Cơ sở lý thuyết

## Activation function

---

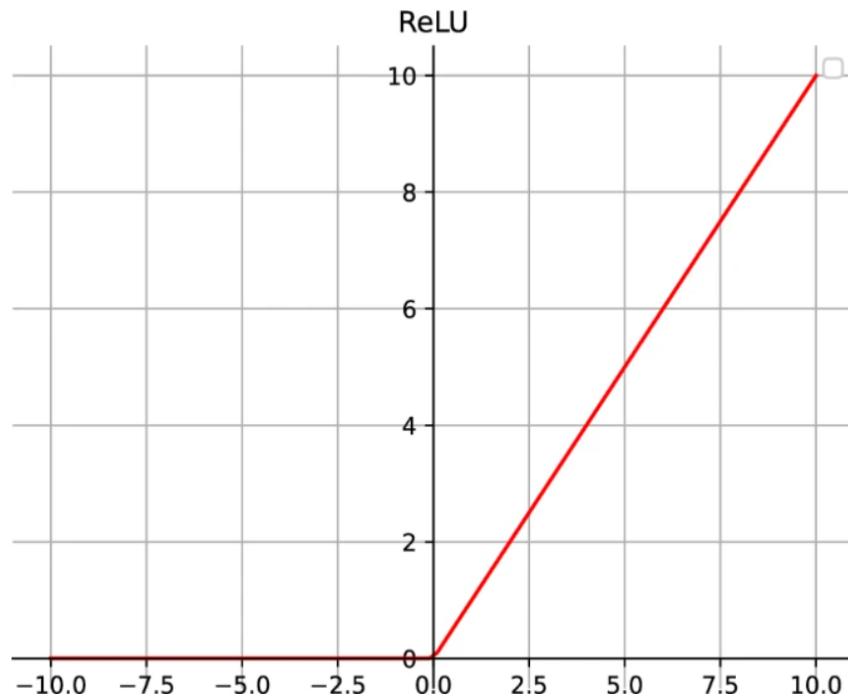
---

- **Hàm kích hoạt (activation function)** là một **phép toán phi tuyến (non-linear function)** được áp dụng **sau mỗi lớp tính toán (như convolution hoặc fully connected)**.
- Biến đổi đầu ra của neuron, giúp mạng **học được các mối quan hệ phức tạp, phi tuyến** trong dữ liệu.
- Một số hàm kích hoạt thường sử dụng : Sigmoid, Tanh, Softmax, Relu và Leaky ReLU.

# Cơ sở lý thuyết

## Activation function

$$f(x) = \max(0, x)$$



- **ReLU** có nhiệm vụ chuyển toàn bộ giá trị âm trong kết quả lấy từ lớp Convolution thành giá trị 0.
- Ý nghĩa :
  - Tăng **tính phi tuyến**, giúp mạng **học được đặc trưng phức tạp** hơn.
  - Giảm hiện tượng **vanishing gradient** so với các hàm sigmoid/tanh.

# Cơ sở lý thuyết

## Pooling layer

0	1	0	7
4	8	2	4
7	6	5	0
3	3	2	9

Max pooling

8	7
7	9

0	1	0	6
4	7	1	1
2	6	5	0
3	3	2	9

Average pooling

Average pooling

3	2
3.5	4

- **Giảm kích thước** feature map, **giữ lại thông tin quan trọng nhất**.
- Giúp **giảm dư thừa**, làm mạng **nhỏ, nhanh và ổn định hơn**.
- Hai kỹ thuật phổ biến :
  - **Max Pooling**: lấy giá trị lớn nhất.
  - **Average Pooling**: lấy giá trị trung bình.

# Cơ sở lý thuyết

## Fully Connected Layer (FC Layer)

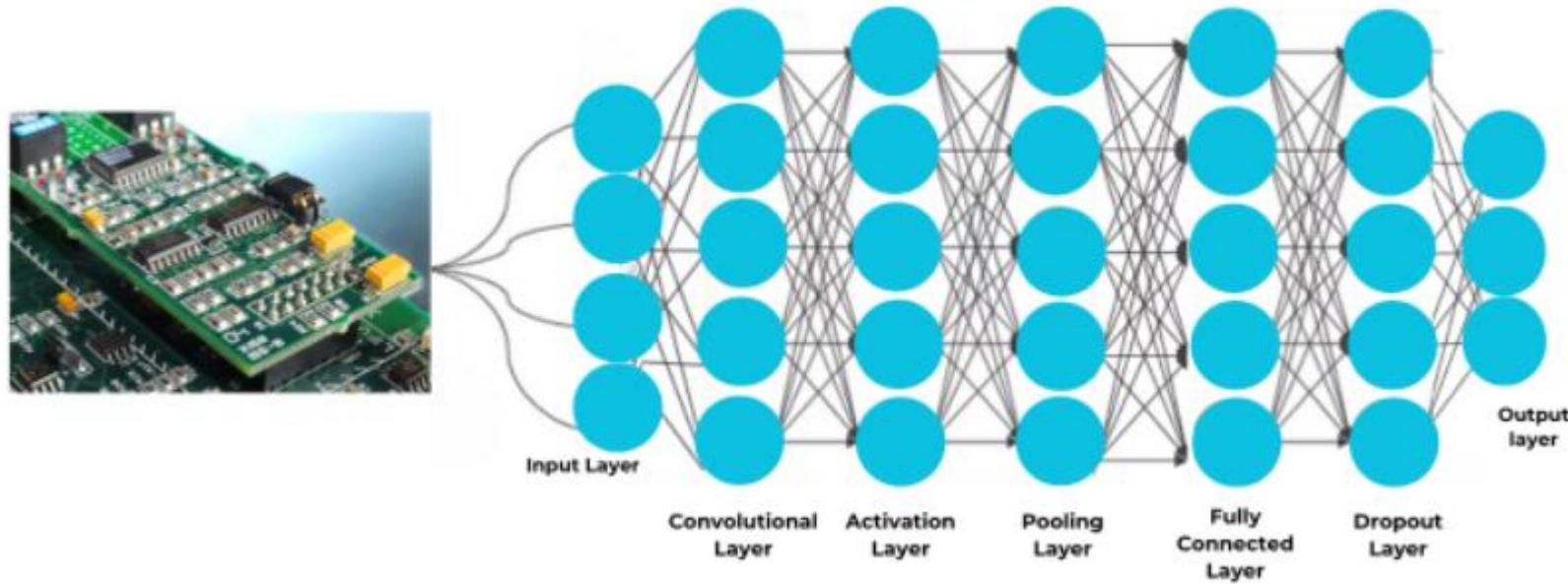
---

---

- Là **lớp kết nối đầy đủ** giữa các neuron
- Nhận đầu vào từ các **feature map** đã được trích xuất.
- Có nhiệm vụ **tổng hợp và phân loại** đặc trưng thành các nhãn đầu ra.
- Giúp **mạng học mối quan hệ phi tuyến** giữa các đặc trưng.

# Cơ sở lý thuyết

## Hoạt động của mô hình CNN



- Được tạo bằng cách **kết nối các layer**: **Convolution** → **ReLU** → **Pooling** → **Fully Connected**.
- **Bắt đầu** với Convolutional Layer để trích đặc trưng.
- Các lớp sau có thể lặp lại **Convolution/Pooling** tùy kiến trúc.
- **Kết thúc** bằng Fully Connected Layer để **phân loại đầu ra**.

>>> Nội dung tiếp theo

## Ứng dụng nhận dạng chữ viết và hình dạng đơn giản

---

---

- Tiền xử lý ảnh đầu vào
- Trích chọn đặc trưng bằng CNN
- Kết hợp các đặc trưng

# Ứng dụng

## Tiền xử lý ảnh đầu vào

---

---

- **Chuẩn hóa kích thước:** Resize về kích thước cố định (28x28 cho MNIST).
- **Kênh màu:** RGB → Grayscale (1 kênh, dùng cho chữ viết tay).

$$Gray = 0.299R + 0.587G + 0.114B$$

- Sử dụng bộ lọc Gauss để lọc ảnh
- Dùng thuật toán Otsu để phân ngưỡng
- Tăng cường dữ liệu (Augmentation):
  - Xoay  $\pm 10\text{--}15^\circ$ , dịch, co giãn, biến dạng đàn hồi.
  - Giúp mô hình học tổng quát, chịu biến đổi tốt.

# Ứng dụng

## Trích chọn đặc trưng bằng CNN

---

---

- Convolution:
  - Dùng *kernel* ( $3 \times 3$ ,  $5 \times 5$ ...) trượt trên ảnh → tạo *feature map*.
  - Mỗi kernel học ra một loại đặc trưng (cạnh, góc,...).
- Hàm kích hoạt (Activation):
  - *ReLU* → thêm phi tuyến, tránh “neuron chết”.
- Pooling :
  - *Max Pooling* ( $2 \times 2$ ) → giảm kích thước, giữ đặc trưng nổi bật.

# Ứng dụng

## Trích chọn đặc trưng bằng CNN

---

---

- **Tầng đầu**: cạnh, góc, nét cơ bản.
  - **Tầng giữa**: motif – nét cong, gạch, phần ký tự.
  - **Tầng cuối**: cấu trúc tổng thể (chữ, hình dạng).
- => CNN tự động học từ **đặc trưng cục bộ** → **tổng quát** (local → global).

# Ứng dụng

## Kết hợp đặc trưng

---

---

- Sử dụng Fully Connected Layer để kết hợp các đặc trưng :
  - **Flatten:** chuyển feature map => vector 1D.
  - **Fully Connected Layer:** kết hợp đặc trưng để dự đoán lớp.
  - **Activation + Dropout (nếu có):** tăng phi tuyến, giảm overfitting.
  - **Output (Softmax):** xuất xác suất cho từng lớp