

Chương 11. Convolutional Neural Network



TRÍ TUỆ NHÂN TẠO *Artificial Intelligence*

Đoàn Vũ Thịnh
Khoa Công nghệ Thông tin
Đại học Nha Trang
Email: thinhhdv@ntu.edu.vn

Nha Trang, 06-2023

Chương 11. Convolutional Neural Network

■ Khái niệm

- Yann LeCun, giám đốc nhóm nghiên cứu AI của Facebook, là người tiên phong xây dựng mạng nơ-ron tích chập đầu tiên có tên LeNet vào năm 1988. LeNet được sử dụng cho các nhiệm vụ nhận dạng ký tự như đọc mã zip và chữ số.
- Mạng nơ ron tích chập hay mạng nơ ron chuyển tiếp nguồn dữ liệu (feed-forward neural network) thường được sử dụng để phân tích hình ảnh trực quan bằng cách xử lý dữ liệu với cấu trúc liên kết giống như lưới. Nó còn được gọi là ConvNet.

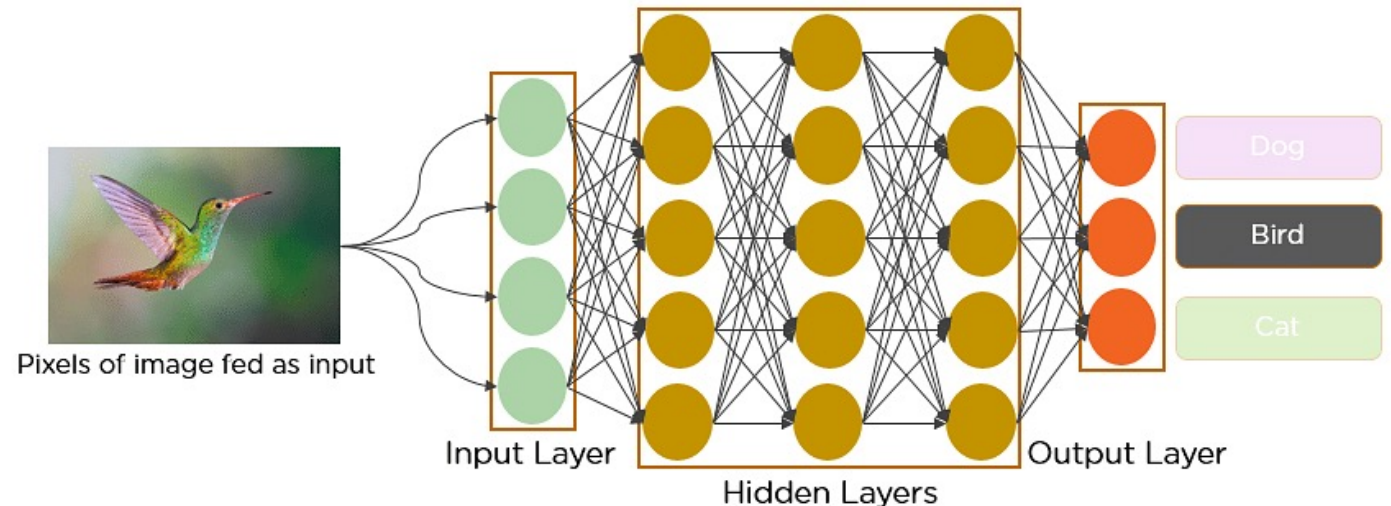


Yann LeCun, 2018

Chương 11. Convolutional Neural Network

■ Khái niệm

- Đầu tiên, cung cấp các pixel của hình ảnh dưới dạng mảng vào lớp đầu vào của mạng nơ ron.
- Thực hiện trích xuất đặc trưng bằng cách thực hiện các phép toán và thao tác khác nhau. Có nhiều lớp ẩn như lớp chập (convolution), lớp ReLU và lớp gộp (pooling).
- Cuối cùng, có một lớp được kết nối đầy đủ (fully connected) để xác định đối tượng trong ảnh.



Chương 11. Convolutional Neural Network

■ Ví dụ

- Trong CNN, hình ảnh được biểu diễn dưới dạng một mảng các giá trị pixel.

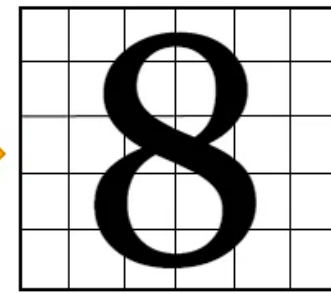
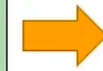
- Phép toán tích chập (convolution) được tính như sau:

- **Ma trận đầu vào (I):**

$$I(m, n) = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$



Real Image of the digit 8



Represented in the form of an array



0	0	1	1	0	0
0	1	0	0	1	0
0	0	1	1	0	0
0	1	0	0	1	0
0	0	1	1	0	0

Digit 8 represented in the form of pixels of 0's and 1's

Ma trận bộ lọc (K):

$$K(p, q) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$F(i, j) = \sum_{u=1}^p \sum_{v=1}^q I(i + u - 1, j + v - 1) \cdot K(u, v)$$

Chương 11. Convolutional Neural Network

■ Phương pháp tính tích chập

$$I = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$F_0 = 3$$

$$K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$F(i, j) = \sum_{u=1}^3 \sum_{v=1}^3 I(i + u - 1, j + v - 1) \cdot K(u, v)$$

■ Ta có ma trận kết quả

$$F_0 = \begin{bmatrix} 0 * 1 + 0 * 0 + 1 * 1 \\ 0 * 0 + 1 * 1 + 0 * 0 \\ 0 * 1 + 0 * 0 + 1 * 1 \end{bmatrix}$$

$$F(0,0) =$$

$$\begin{aligned} & I(0,0) \cdot K(1,1) + I(0,1) \cdot K(1,2) + I(0,2) \cdot K(1,3) \\ & + I(1,0) \cdot K(2,1) + I(1,1) \cdot K(2,2) + I(1,2) \cdot K(2,3) \\ & + I(2,0) \cdot K(3,1) + I(2,1) \cdot K(3,2) + I(2,2) \cdot K(3,3) \end{aligned}$$

$$F_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}^+$$

Chương 11. Convolutional Neural Network

■ Phương pháp tính tích chập

$$I = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} 3 & 2 & 2 & 3 \\ 0 & 3 & 3 & 0 \\ 3 & 2 & 2 & 3 \end{bmatrix}$$

$$K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

$$I = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$F = \begin{bmatrix} 3 & & & & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

$$K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

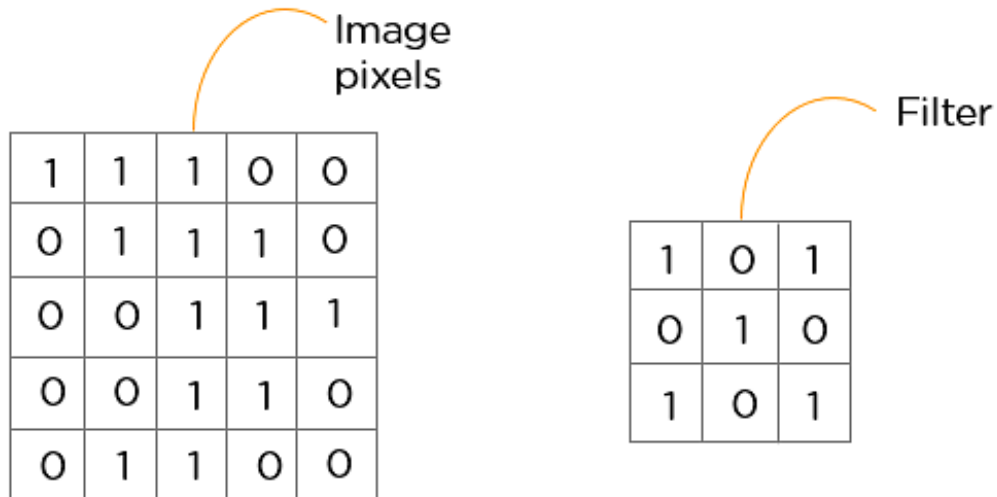
Chương 11. Convolutional Neural Network

- Các lớp bên trong CNN
- Mạng nơ ron tích chập có nhiều lớp ẩn giúp trích xuất thông tin từ hình ảnh. Bốn lớp quan trọng trong CNN là:
- Convolution layer
- ReLU layer
- Pooling layer
- Fully connected layer

Chương 11. Convolutional Neural Network

■ Convolution layer

- Đây là bước đầu tiên trong quá trình trích xuất các đặc điểm có giá trị từ một hình ảnh. Một lớp tích chập có một số bộ lọc thực hiện thao tác tích chập. Mỗi hình ảnh được coi là một ma trận các giá trị pixel.



<https://www.simplilearn.com>

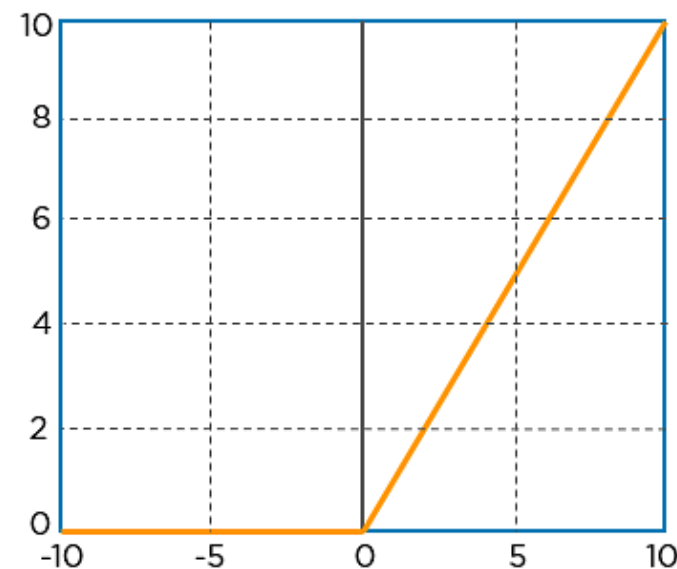
- Ví dụ, ảnh 5x5 có giá trị pixel là 0 hoặc 1. Ngoài ra còn có ma trận bộ lọc có kích thước 3x3. Trượt ma trận bộ lọc lên hình ảnh và tính tổng của tích các điểm để có được ma trận tích chập.

Chương 11. Convolutional Neural Network

■ ReLU

- ReLU (rectified linear unit) là viết tắt của đơn vị tuyến tính được chỉnh lưu. Sau khi trích xuất các bản đồ đặc trưng, bước tiếp theo là chuyển chúng sang ReLU.
- ReLU thực hiện thao tác theo từng phần tử và đặt tất cả các pixel âm thành 0. $f(x) =$

$$\max(0, x) \text{ hay } f(x) = \begin{cases} 0, & \text{nếu } x \leq 0 \\ x, & \text{nếu } x > 0 \end{cases}$$

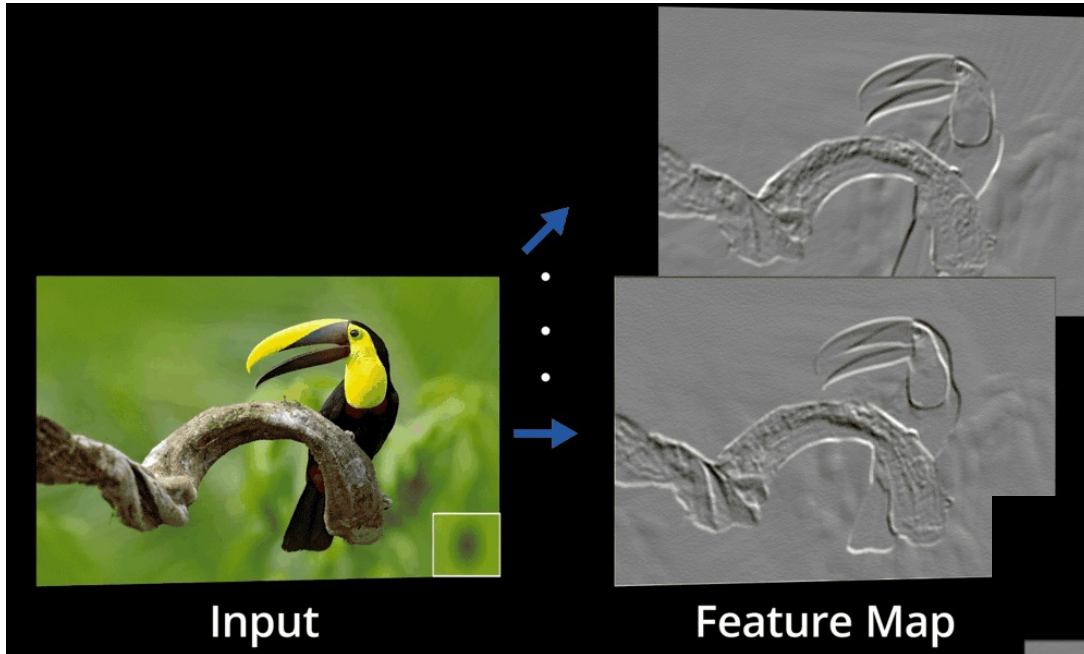


$$R(z) = \max(0, z)$$

<https://www.simplilearn.com>

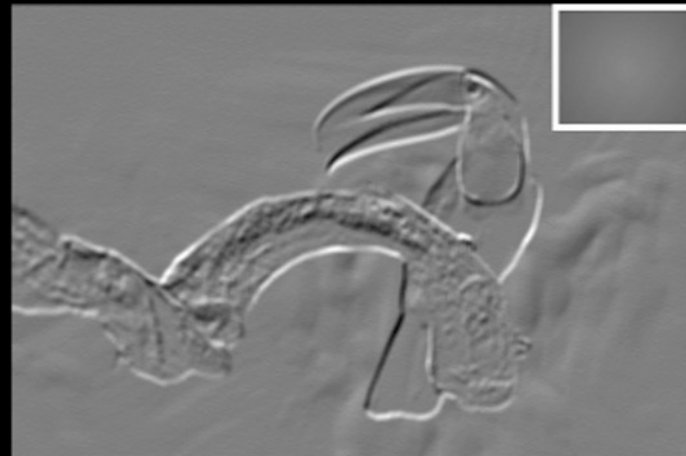
Chương 11. Convolutional Neural Network

- ReLU layer



<https://www.simplilearn.com>

Input Feature Map



Chương 11. Convolutional Neural Network

■ Pooling layer

- Lớp Pooling được sử dụng để giảm kích thước của dữ liệu và giảm độ phức tạp của mô hình, thường được áp dụng sau các lớp convolution để giảm số lượng tham số và tính toán. Các phép toán phổ biến nhất là Max Pooling và Average Pooling.
- Pooling layer có tác dụng: Giảm kích thước của dữ liệu, Giảm overfitting và Tăng tính chống chéo (có khả năng nhận biết đặc trưng tốt hơn ở các vị trí không chính xác)

<https://www.simplilearn.com>

Rectified feature map

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

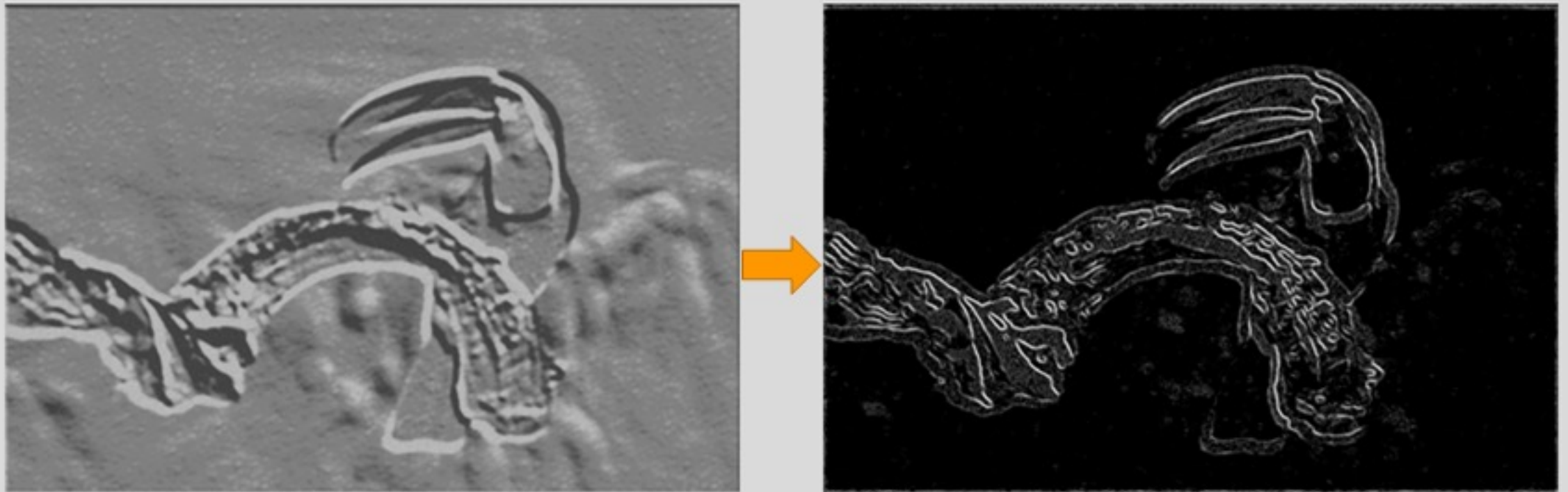
$$\text{Max}(3, 4, 1, 2) = 4$$

Pooled feature map

6	8
4	7

Chương 11. Convolutional Neural Network

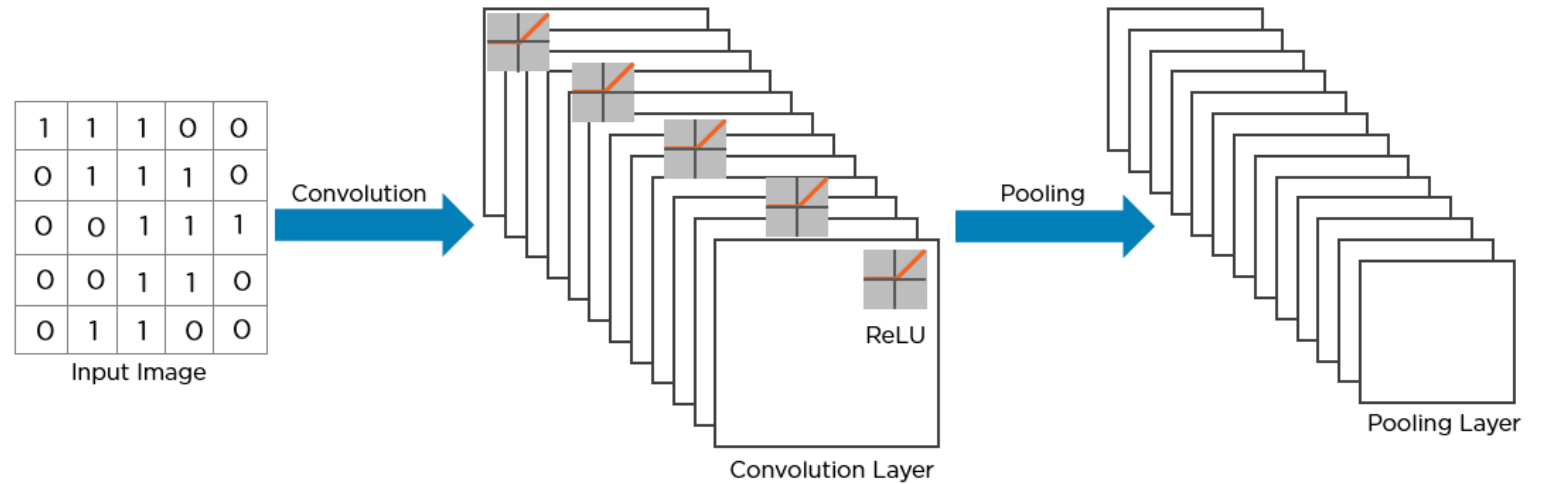
- Pooling layer
- Lớp Pooling sử dụng nhiều bộ lọc khác nhau để xác định các phần khác nhau của hình ảnh như cạnh, góc, thân, lông, mắt và mỏ.



Chương 11. Convolutional Neural Network

■ Các lớp bên trong CNN

- Convolution layer
- ReLU
- Pooling layer
- Fully connected layer



Chương 11. Convolutional Neural Network

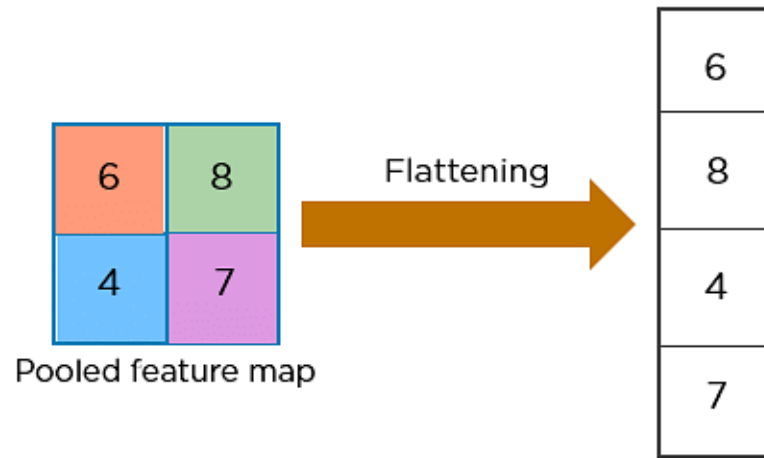
■ Các lớp bên trong CNN

- Convolution layer

- ReLU layer

- Pooling layer

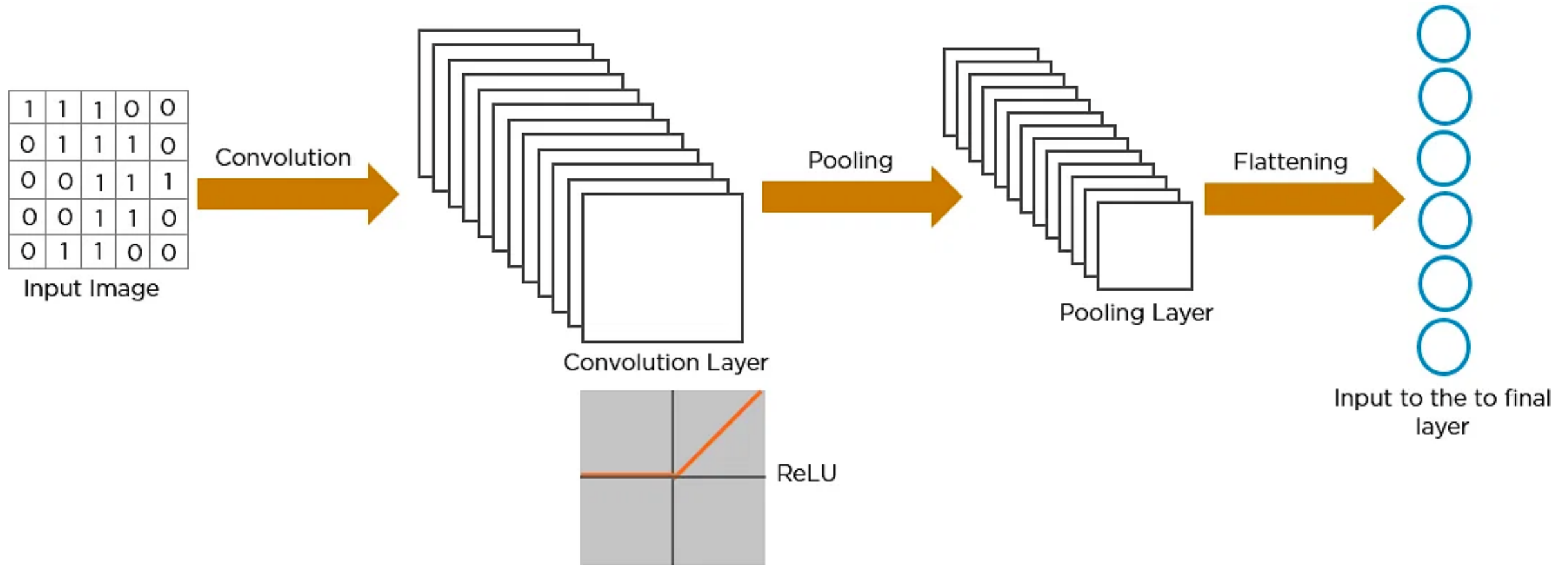
- Fully connected layer



- Bước tiếp theo trong quy trình này được gọi là làm phẳng (Flattening). Làm phẳng được sử dụng để chuyển đổi tất cả các mảng 2 chiều tổng hợp từ các feature map sau bước pooling thành một vector tuyến tính duy nhất.
- Ma trận phẳng được đưa vào làm đầu vào cho lớp kết nối đầy đủ (fully connected layer) để phân loại hình ảnh.

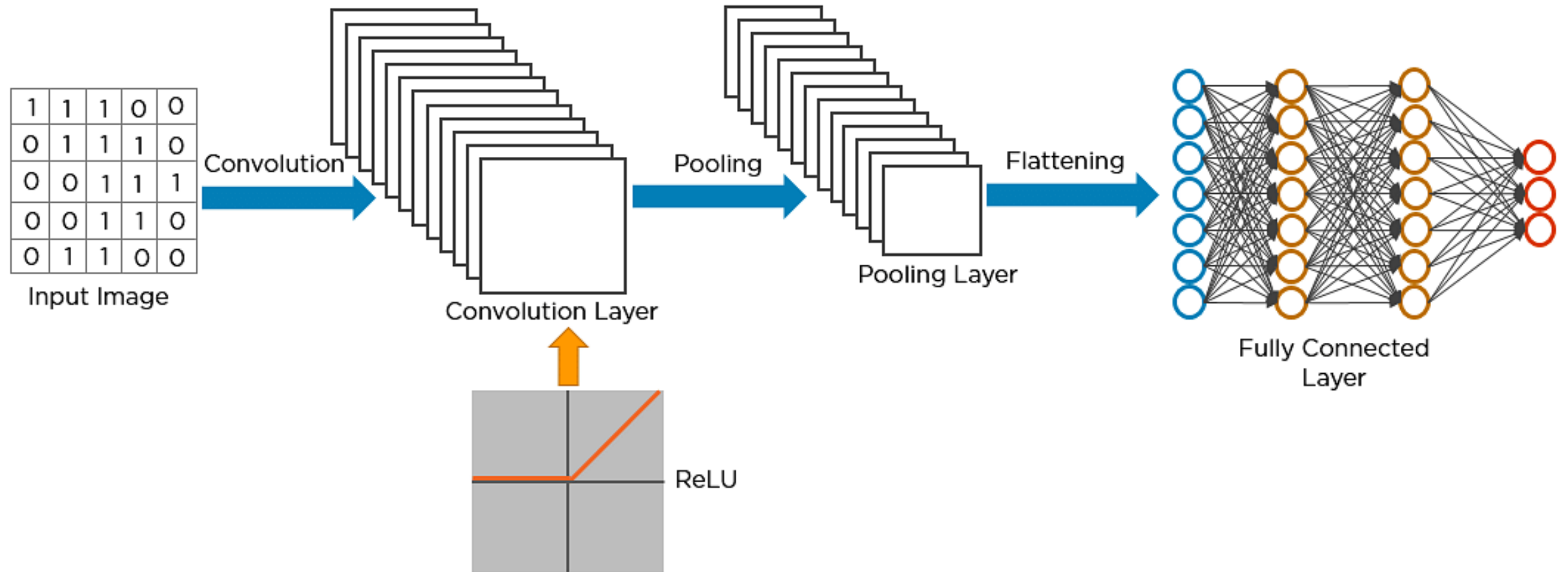
Chương 11. Convolutional Neural Network

■ Fully connected layer



Chương 11. Convolutional Neural Network

■ Fully connected layer

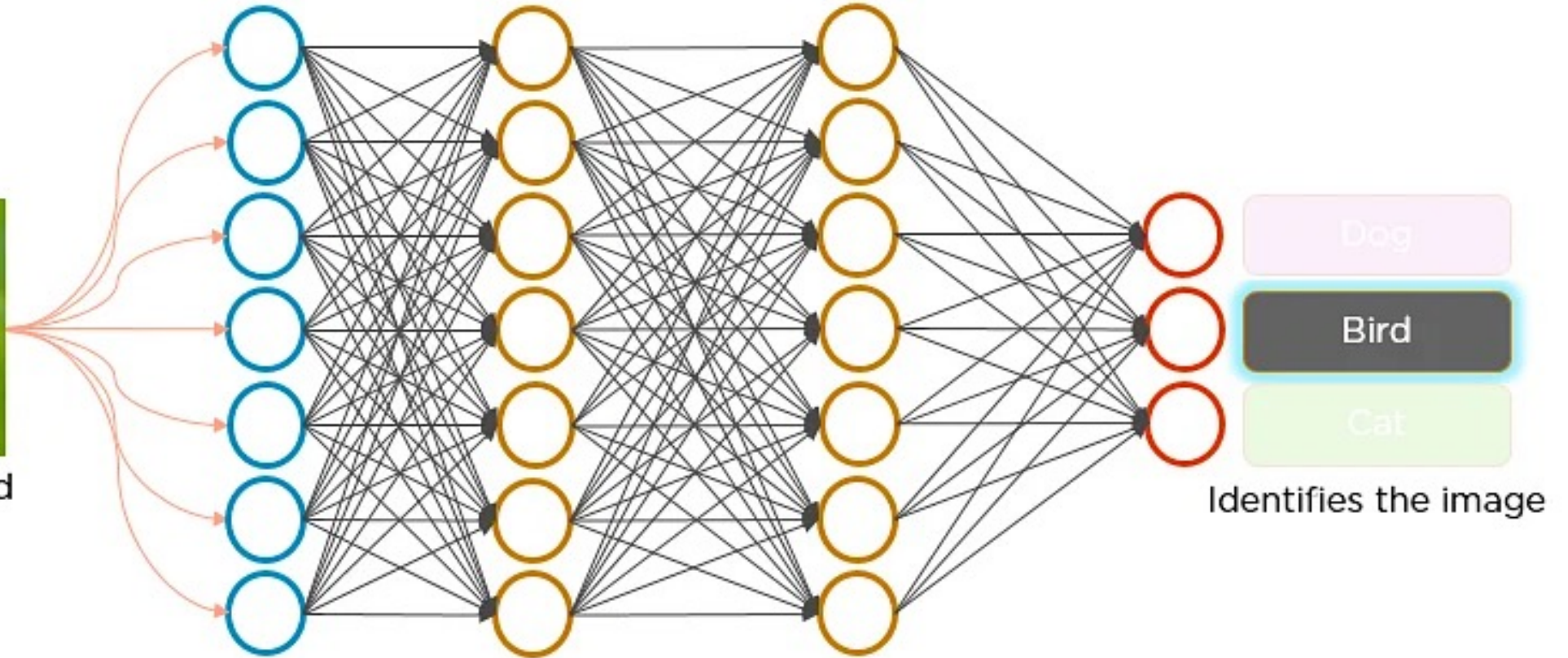


Chương 11. Convolutional Neural Network

- Fully connected layer



Pixels from the flattened matrix fed as input



Chương 11. Convolutional Neural Network

- Demo in Python