TRƯỜNG ĐẠI HỌC TRÀ VINH

TRƯỜNG KỸ THUẬT VÀ CÔNG NGHỆ

KHOA CÔNG NGHỆ THÔNG TIN



BÀI TẬP DỰ ÁN: Tìm hiểu về mô hình CNN trong thị giác máy tính và ứng dụng trong nhận dạng khuôn mặt

NHÓM THỰC HIỆN : NHÓM 1

Bao Gồm Các Thành Viên :

Lê Hoàng Ân 119525020

Đặng Minh Duy 119525002

Nguyễn Hoàng Gia Minh 119525010

Vĩnh Long , 10/2025

Mục Lục

[I. Khái quát về mạng CNN 3](#_Toc1703979772)

[1. Khái niệm về mạng CNN 3](#_Toc1380726734)

[3. Kĩ Thuật CNN: 3](#_Toc1087798998)

[4. Ưu điểm và nhược điểm của CNN 3](#_Toc1091349777)

[6. Bộ lọc(kernel/filter): 5](#_Toc86440988)

[7. Bước di chuyển(stride) 5](#_Toc2027285886)

[8.Đệm(padding): 6](#_Toc264671576)

[9. Lớp gộp – Pooling Layer 6](#_Toc382155213)

[9.1 Max Pooling (Gộp cực đại) : 7](#_Toc104399555)

[9.2 Average Pooling (Gộp trung bình) : 7](#_Toc1910007886)

[9.3 Sum Pooling (Gộp tổng) : 7](#_Toc1280339325)

[9.4 Fully connected layer (Tầng kết nối đầy đủ) : 8](#_Toc1435344332)

[9.5 Weights and Biases (Trọng số và độ chệch) 8](#_Toc1361818045)

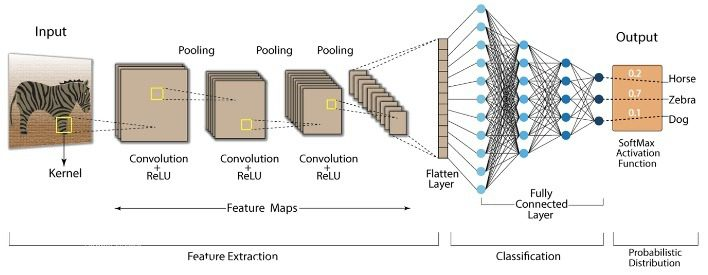
[9.6 Activation Function (Hàm kích hoạt) 8](#_Toc778995637)

CHƯƠNG 1 :TÌM HIỂU CHUNG VỀ MẠNG CNN

## 1. Khái niệm về mạng CNN

- CNN là một mô hình mạng Neural tích chập (CNN) là 1 trong những mô

hình để nhận dạng và phân loại hình ảnh. Trong đó, xác định đối tượng và nhận dạng khuôn mặt là 1 trong số những lĩnh vực mà CNN được sử dụng rộng rãi.



2.CNN phân loại hình ảnh bằng cách nào?

- CNN phân loại hình ảnh bằng cách lấy 1 hình ảnh đầu vào, xử lý và phân loại nó theo các hạng mục nhất định (Ví dụ: Chó, Mèo, Hổ, ...). Máy tính coi hình ảnh đầu vào là 1 mảng pixel và nó phụ thuộc vào độ phân giải của hình ảnh. Dựa trên độ phân giải hình ảnh, máy tính sẽ thấy H x W x D (H: Chiều cao, W: Chiều rộng, D: Độ dày).

//hình rgb

## 3. Kĩ Thuật CNN:

-Về kỹ thuật, mô hình CNN để training và kiểm tra, mỗi hình ảnh đầu vào sẽ chuyển nó qua 1 loạt các lớp tích chập với các bộ lọc (Kernals), tổng hợp lại các lớp được kết nối đầy đủ (Full Connected) và áp dụng hàm Softmax để phân loại đối tượng có giá trị xác suất giữa 0 và 1.

## 4. Ưu điểm và nhược điểm của CNN

**\*Ưu điểm**

***- Trích xuất đặc trưng tự động:***CNN tự động học cách trích xuất các đặc trưng quan trọng từ dữ liệu ảnh mà không cần phải xác định rõ ràng các đặc trưng cụ thể.

***- Xử lý dữ liệu không gian:*** CNN được thiết kế đặc biệt để xử lý dữ liệu không gian như ảnh và âm thanh. Các lớp tích chập và gộp cho phép nó hiệu quả trong việc trích xuất thông tin từ các cấu trúc không gian này.

***- Phân loại và nhận dạng tốt:*** CNN đã đạt được hiệu suất ấn tượng trong các nhiệm vụ phân loại và nhận dạng. Điều này bao gồm nhận biết khuôn mặt, phân loại đối tượng, dự đoán từ ảnh, và nhiều nhiệm vụ khác.

***- Tích hợp thông tin cục bộ:*** Các lớp tích chập trong CNN cho phép mạng tập trung vào việc xử lý thông tin cục bộ trong các phần nhỏ của dữ liệu ảnh. Điều này giúp mạng nhận biết các đặc trưng cục bộ và tạo ra biểu diễn chất lượng cao.

***- Hiệu suất cao:*** Các mô hình CNN tiên tiến có khả năng đạt được hiệu suất rất cao trong các thách thức như các cuộc thi nhận dạng hình ảnh (ImageNet) và các ứng dụng thực tiễn như xe tự hành

**\*Nhược điểm**

***- Số lượng tham số và tài nguyên tính toán:*** Các mạng CNN tiên tiến thường có hàng triệu hoặc thậm chí hàng tỷ tham số. Điều này đòi hỏi tài nguyên tính toán lớn để huấn luyện và triển khai, đặc biệt là trên các thiết bị có tài nguyên hạn chế như điện thoại di động.

***- Dữ liệu huấn luyện hạn chế:*** Mạng CNN yêu cầu một lượng lớn dữ liệu huấn luyện để học hiệu quả. Nếu dữ liệu không đủ đa dạng hoặc không đại diện, mô hình có thể không học được các biểu diễn tổng quát.

**- *Khả năng biểu diễn không tốt cho dữ liệu không gian tương đối dài:*** Mạng CNN không thể hiệu quả biểu diễn thông tin về thứ tự và tương quan không gian dài hơn, chẳng hạn như các dãy thời gian dài

**- *Chuẩn bị dữ liệu:*** Trước khi áp dụng CNN, dữ liệu thường cần phải được chuẩn bị và tiền xử lý một cách cẩn thận, bao gồm việc thay đổi kích thước ảnh, chuẩn hóa dữ liệu và loại bỏ nhiễu.

5. Lớp tích chập - Convolution Layer

-Trong CNN (Convolutional Neural Network), **lớp tích chập** là lớp dùng để trích xuất đặc trưng từ ảnh, ví dụ: cạnh, góc, hoa văn, vật

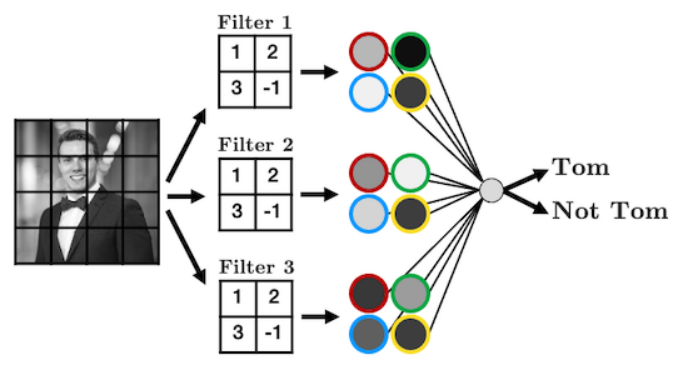
Thể thay vì nhìn toàn bộ ảnh một lúc, nó dùng bộ lọc( kernel/filter**)** (ma trận nhỏ, như 3×3, 5×5) để quét từng vùng nhỏ của ảnh và tạo ra **feature map**.

## 6. Bộ lọc(kernel/filter):

-Đây là ma trận nhỏ chứa các trọng số được áp dụng lên vùng cục bộ của dữ liệu đầu vào

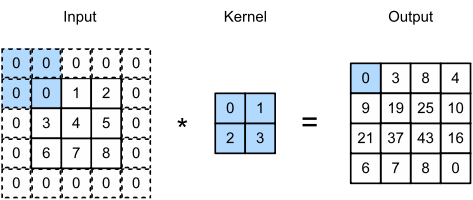
(ví dụ: hình ảnh). Bộ lọc này có thể là một ma trận có kích thước nhỏ,

chẳng hạn 3x3 hoặc 5x5.

**D**

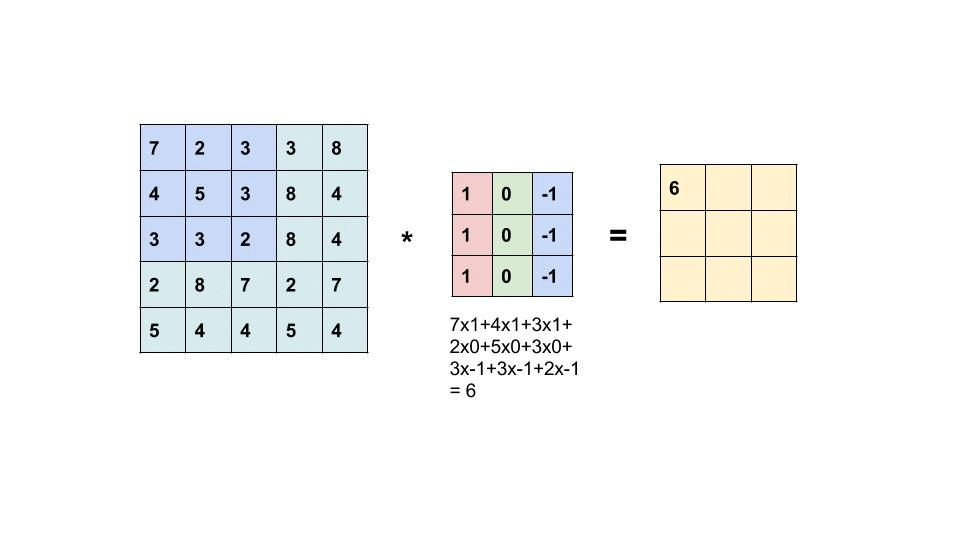
## 7. Bước di chuyển(stride)

-Stride là số pixel thay đổi trên ma trận đầu vào. Khi stride là 1 thì ta di chuyển các kernel 1 pixel. Khi stride là 2 thì ta di chuyển các kernel đi 2 pixel và tiếp tục như vậy. Hình dưới là lớp tích chập hoạt động với stride là 2.



## 8.Đệm(padding):

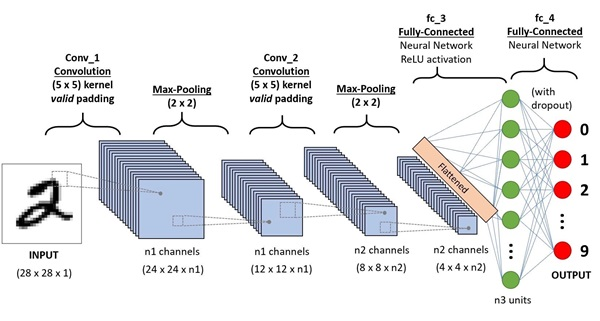
- Đệm là việc thêm các giá trị 0 vào xung quanh dữ liệu đầu vào trước khi áp dụng phép tích chập. Điều này giúp duy trì kích thước của dữ liệu sau khi đi qua lớp tích chập.



## 9. Lớp gộp – Pooling Layer

***\*Khái niệm***

- Lớp gộp (Pooling Layer) là một phần quan trọng trong kiến trúc của mạng nơ-ron sâu (deep neural networks), đặc biệt là trong các mạng nơ-ron tích chập (convolutional neural networks – CNN), được sử dụng trong xử lý hình ảnh và các tác vụ liên quan đến dữ liệu lưới như dữ liệu âm thanh.



### 9.1 Max Pooling (Gộp cực đại) :

-Max pooling là một loại phép gộp được sử dụng trong mạng nơ-ron tích chập (CNN) và các kiến trúc liên quan. Nó thường được áp dụng sau lớp tích chập để giảm kích thước của đầu vào và tạo ra các phiên bản đã tạo ra của đặc trưng cục bộ.

### 9.2 Average Pooling (Gộp trung bình) :

***\*Khái niệm***

-Average Pooling là một phép tích chập thường được sử dụng trong mạng học sâu và xử lý ảnh. Nó là một phần quan trọng của các kiến trúc mạng thần kinh như Convolutional Neural Networks (CNNs) để giảm kích thước của đặc trưng (feature maps) trong quá trình trích xuất thông tin từ hình ảnh.

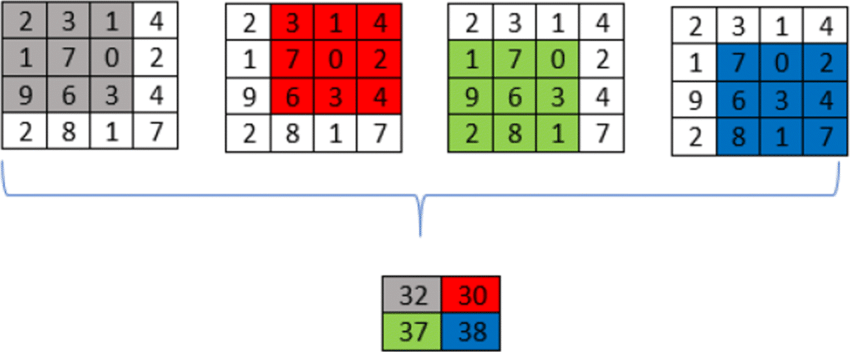
***\* Ưu diểm chính***

-Ưu điểm chính của Average Pooling là nó giúp giảm kích thước của dữ liệu mà vẫn giữ lại thông tin quan trọng. Điều này có thể giúp làm giảm lượng tính toán và giảm overfitting trong quá trình huấn luyện mạng.

### 9.3 Sum Pooling (Gộp tổng) :

-Sum Pooling là một phương pháp trong xử lý ảnh và mạng học sâu, được sử dụng để giảm kích thước của feature maps trong quá trình trích xuất thông tin từ hình ảnh.

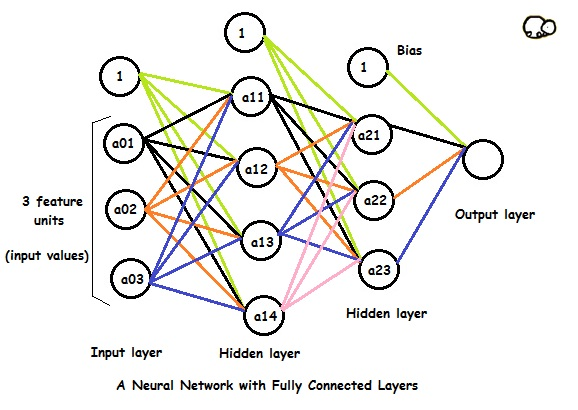
-Tương tự như Average Pooling và Max Pooling, Sum Pooling cũng là một loại phép gộp thông tin trong mạng neural. Mục tiêu của Sum Pooling tương tự như các phương pháp gộp khác, đó là giảm kích thước của dữ liệu trong quá trình huấn luyện mạng, giảm lượng tính toán và có thể giúp kiểm soát overfitting.



### 9.4 Fully connected layer (Tầng kết nối đầy đủ) :

***\*Khái niệm***

-Fully connected layers (còn gọi là dense layers) là một loại tầng trong kiến trúc mạng neural, bao gồm các neuron được kết nối với tất cả các neuron ở tầng trước và tầng sau. Đây là tầng mạng neural truyền thống, trong đó mỗi neuron nhận đầu vào từ tất cả các neuron ở tầng trước đó và truyền đầu ra cho tất cả các neuron ở tầng sau.



***\*Cách hoạt động của Fully Connected Layer***

### 9.5 Flattening (Phẳng hóa)

Trong ngữ cảnh của Convolutional Neural Networks (CNN), đầu ra từ các tầng convolutional và pooling layers thường là một tensor 3D hoặc 4D. Để đưa chúng vào fully connected layers, cần phải làm phẳng (flatten) chúng thành một vector 1D. Điều này thường được thực hiện trước khi các fully connected layers.

#### 9.6 Weights and Biases (Trọng số và độ chệch)

-Mỗi neuron trong fully connected layers có một trọng số (weight) tương ứng với mỗi kết nối đến neuron trong tầng trước đó. Ngoài ra, mỗi neuron còn có một bias. Các trọng số và bias này là các tham số mạng neural được học để tối ưu hóa hiệu suất mạng.

#### 9.7 Activation Function (Hàm kích hoạt)

- Sau khi tính tổng trọng số đầu vào từ các neuron ở tầng trước kết hợp với bias, đầu ra của neuron thường được đưa qua một hàm kích hoạt phi tuyến (như ReLU, Sigmoid, hoặc Tanh) để tạo ra đầu ra của tầng. Hàm kích hoạt này thêm tính phi tuyến tính vào tầng fully connected.