**TRƯỜNG ĐẠI HỌC TRÀ VINH**

**TRƯỜNG KỸ THUẬT VÀ CÔNG NGHỆ**

**KHOA CÔNG NGHỆ VÀ THÔNG TIN**



BÀI TẬP DỰ ÁN

TÌM HIỂU VỀ MÔ HÌNH CNN TRONG THỊ GIÁC MÁY TÍNH VÀ ỨNG DỤNG TRONG NHẬN DIỆN KHUÔN MẶT.

NHÓM 1 LỚP DA25TTNT

LÊ HOÀNG ÂN 119525020

ĐẶNG MINH DUY 119525002

NGUYỄN HOÀNG GIA MINH 119525010

VĨNH LONG , 10,2025

MỤC LỤC

[CHƯƠNG I : TÌM HIỂU CHUNG VỀ MẠNG CNN 4](#_Toc212582710)

[I.TỔNG QUAN VỀ MANG CNN 4](#_Toc212582711)

DANH MỤC HÌNH ẢNH

DANH MỤC TÀI LIỆU THAM KHẢO

https://tasdigital.net/pooling-la-gi.html

# CHƯƠNG I : TÌM HIỂU CHUNG VỀ MẠNG CNN

## I.TỔNG QUAN VỀ MANG CNN

Công nghệ AI (trí tuệ nhân tạo) là lĩnh vực nghiên cứu và phát triển các hệ thống máy tính có khả năng thực hiện các nhiệm vụ mà thường cần đến trí thông minh của con người. AI bao gồm nhiều công nghệ và kỹ thuật khác nhau, như: Machine Learning (Học máy), Natural Language Processing (Xử lý ngôn ngữ tự nhiên), Computer Vision (Thị giác máy tính), Robotics (Ngành robot) . Và CNN cũng thuộc Deep learning và là một lĩnh vực con của Machine learing nơi các hệ thống máy tính học và cải thiện khả năng của mình thông qua các thuật toán phức tạp, lấy cảm hứng từ cách hoạt động của não bộ con người.

Convolutional Neural Network là một loại mô hình mạng nơ-ron nhân tạo được sử dụng rộng rãi trong lĩnh vực thị giác máy tính và xử lý ảnh. CNN được thiết kế để nhận dạng và phân loại hình ảnh dựa trên cấu trúc của nó.

1. Khái niệm của mạng Nơ-Ron trong thị giác máy tính CNN.

1.1 Khái quát chung

-Nơ ron nhân tạo có các mô đun phần mềm, được gọi là nút. Mạng nơ ron nhân tạo là chương trình phần mềm hoặc thuật toán mà về cơ bản, sử dụng hệ thống máy tính để giải quyết phép toán.

**1. Thị giác máy tính (Computer Vision)**

Nhìn vào một bức ảnh, một người với thị giác bình thường có thể dễ dàng mô tả nội dung, nhận biết và phát hiện các đối tượng được thể hiện trong bức ảnh cũng như vị trí chính xác của chúng. Tuy nhiên, việc này (đọc và hiểu một bức ảnh) khó khăn hơn nhiều đối với máy tính khi “nó” “nhìn” mỗi bức ảnh chỉ đơn thuần là một ma trận số (tập hợp các điểm ảnh - **pixel** biểu diễn dưới dạng số theo một hệ cụ thể thường là RGB (Red - Green - Blue)). Mục tiêu chính của Thị giác máy tính (Computer Vision) - một nhánh của trí tuệ nhân tạo (Artificial Intelligence) là tìm ra cầu nối giữa ma trận số này và thông tin ngữ nghĩa ẩn chứa trong ảnh. Thị giác máy tính tập trung giải quyết những bài toán như:

* **Phân loại ảnh, miêu tả ảnh**,
* **Phát hiện vật thể trong ảnh**: Xe, con người, đèn giao thông, etc.
* **Tạo ảnh với những phong cách khác nhau**: Hiển thị nội dung ngữ nghĩa của ảnh gốc theo những phong cách khác nhau.

  1.2.CNN phân loại hình ảnh bằng cách nào?

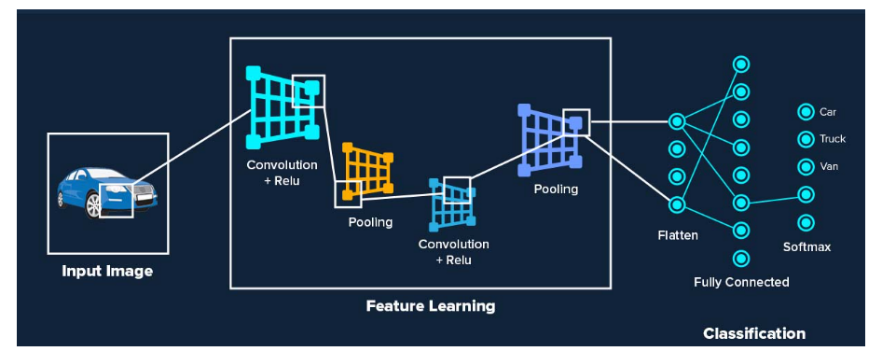
- CNN phân loại hình ảnh bằng cách lấy 1 hình ảnh đầu vào, xử lý và phân loại nó theo các hạng mục nhất định (Ví dụ: Chó, Mèo, Hổ, ...). Máy tính coi hình ảnh đầu vào là 1 mảng pixel và nó phụ thuộc vào độ phân giải của hình ảnh. Dựa trên độ phân giải hình ảnh, máy tính sẽ thấy H x W x D (H: Chiều cao,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

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.

1.1.4Thuật toán CNN:

**Giới thiệu về thuật toán CNN:**

[**Thuật toán CNN**](https://lanit.com.vn/thuat-toan-cnn.html) là một phần quan trọng của Deep Learning – tập hợp các thuật toán nhằm xây dựng mô hình dữ liệu trừu tượng thông qua việc sử dụng nhiều lớp xử lý cấu trúc phức tạp.

CNN hoạt động dựa trên dữ liệu ảnh và được sử dụng rộng rãi trong các ứng dụng như nhận dạng hình ảnh và khuôn mặt. Thuật toán CNN hiện được tích hợp vào các nền tảng mạng phổ biến như Facebook và Google,… Thuật toán CNN nhận đầu vào dưới dạng một mảng hai chiều và xử lý trực tiếp trên hình ảnh, giúp nó trích xuất các đặc trưng quan trọng từ dữ liệu hình ảnh một cách hiệu quả.

Về phần kỹ thuật, thuật toán CNN là mô hình được sử dụng trong quá trình training và kiểm tra dữ liệu hình ảnh. Mỗi hình ảnh đầu vào sẽ đi qua một chuỗi các bước, bao gồm các lớp tích chập với các bộ lọc (Kernel), sau đó là lớp kết nối đầy đủ và cuối cùng áp dụng hàm Softmax để phân loại đối tượng. Kết quả phân loại sẽ là một giá trị xác suất nằm trong khoảng từ 0 đến 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 Nguyên tắc hoạt động (CNN - Convolutional Neural Network).

Đầu vào (Input): Mạng CNN nhận vào dữ liệu đầu vào là các hình ảnh hoặc dữ liệu có cấu trúc dạng lưới khác như âm thanh, video.

*1. Convolution Layer (Lớp tích chập)*

Convolution Layer là lớp đầu tiên để lấy các đặc điểm từ hình ảnh ban đầu. Nó giữ lại các mối quan hệ giữa các pixel bằng cách xem xét các tính năng của hình ảnh thông qua việc sử dụng các ô vuông nhỏ từ dữ liệu gốc. Đây là một phép tính với hai thành phần đầu vào, là ma trận hình ảnh ban đầu và một bộ lọc hoặc hạt nhân.

2. 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

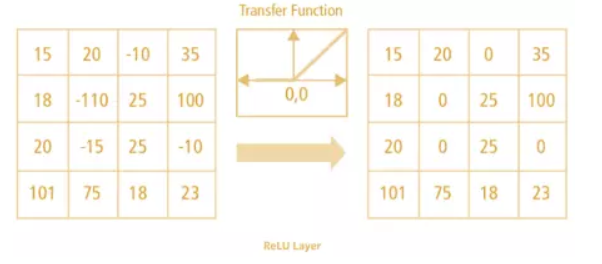
(ví dụ: hình ảnh). Bộ lọc này có thể là một ma trận có kích thước nhỏ,

chẳng hạn 3x3 hoặc 5x5.

*3. Stride (Bước nhảy)*

Stride chính là số pixel luôn thay đổi trên ma trận đầu vào. Khi stride là 1, ta di chuyển kernel 1 pixel, nếu stride là 2 thì di chuyển kernel 2 pixel,…

4.Hàm kích hoạt (Activation):

Hàm ReLU giúp mô hình học những đặc trưng chưa được tìm thấy trong dữ liệu và làm cho quá trình huấn luyện mô hình trở nên hiệu quả hơn.

Một số hàm phi tuyến khác có thể kể đến như: sigmoid, tanh cũng có thể thay thế cho ReLU nhưng ReLU thường được ưu tiên sử dụng bởi hiệu suất tốt

1. Lớp gộp (Pooling):

Bước gộp giúp giảm kích thước của feature map bằng cách thực hiện các phép tổng hợp trên các vùng cục bộ của feature map. Điều này giúp giảm lượng tham số trong mô hình và làm giảm độ phức tạp của quá trình học, pooling có thể có nhiều loại khác nhau:

Max Pooling

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.

**Ưu điểm:** Max pooling rất hiệu quả trong việc phát hiện và giữ lại các đặc trưng nổi bật nhất, sáng nhất như cạnh, góc của vật thể. Nó hoạt động dựa trên giả định rằng đặc trưng càng mạnh (giá trị càng lớn) thì càng quan trọng

 Average Pooling

-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 điểm:** Average Pooling cung cấp một cái nhìn tổng quan, làm mịn các đặc trưng trong một vùng. Nó hữu ích khi cần một biểu diễn tổng quát hơn là chỉ tập trung vào các điểm nổi bật nhất.

6. Bước lặp lại (Repeat):

Các bước tích chập, kích hoạt và gộp có thể được lặp lại nhiều lần để tạo ra các layer ẩn sâu trong mạng.

7 Lớp kết nối đầy đủ (Fully Connected – Dense)

Sau khi các layer tích chập và gộp, các feature map được duỗi thành một vector và được đưa vào các lớp nơ-ron đầy đủ (fully connected) để thực hiện việc phân loại hoặc dự đoán.

Đầu ra (Output)

Cuối cùng, một lớp đầu ra được sử dụng để tạo ra dự đoán cho bài toán cụ thể, như phân loại hình ảnh thành các lớp khác nhau hoặc dự đoán các giá trị trong trường hợp của học có giám sát

***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)*

A diagram of a diagram of a diagram

AI-generated content may be incorrect. - 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.

.