

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

BỘ MÔN THỰC TẬP CƠ SỞ



THỰC TẬP CƠ SỞ

Giảng viên hướng dẫn	: KIM NGỌC BÁCH
Họ và tên sinh viên	: NGUYỄN XUÂN HẢI
Mã sinh viên	: B22DCCN271
Lớp	: D22CQCN07-B
Nhóm	: 13

Báo Cáo Hàng Tuần Lần 1 ngày (8-15/3/2025)

Hà Nội – 2025

BÁO CÁO TIỀN ĐỘ HÀNG TUẦN

1. Mục Tiêu

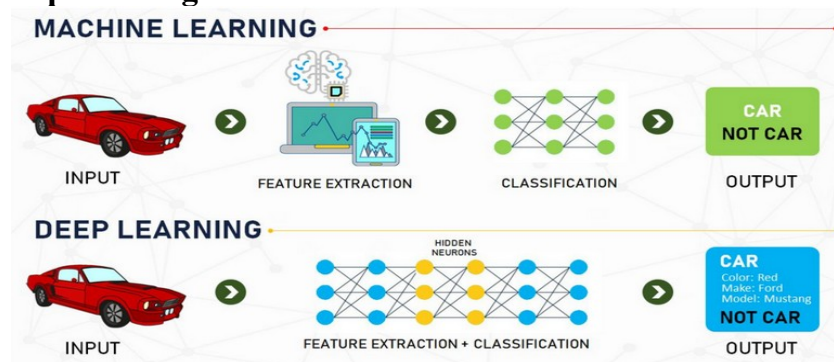
- Mong muốn áp dụng những kiến thức được học vào đề tài
- Phát triển thêm kỹ năng và kiến thức về Computer Vision trong quá trình làm bài thực tập cơ sở

2. Đề tài dự kiến

- Lập trình AI để phân biệt ảnh động vật (6 loại)
- Công nghệ sử dụng: sklearn, Pytorch, torchvision, mạng CNN

3. Báo cáo tiến độ

a, Tìm hiểu về deep learning



Deep Learning (học sâu) cũng có thể được coi là một lĩnh vực thuộc **Machine Learning** – nơi máy tính tự học và cải thiện thông qua các thuật toán. Thế nhưng, Deep Learning được xây dựng dựa trên những khái niệm phức tạp hơn. Mạng lưới thần kinh sâu (DNN - Deep neural networks) bao gồm nhiều lớp tế bào thần kinh khác nhau, có khả năng thực hiện các tính toán rất phức tạp.

b, Cách thức hoạt động

Một mạng nơ-ron bao gồm nhiều lớp (layer) khác nhau, càng nhiều lớp thì mạng sẽ càng “sâu”. Trong mỗi lớp là các node (nút mạng) và được liên kết với các lớp khác bên cạnh.

- Mỗi kết nối giữa các nút sẽ có một trọng số tương ứng, trọng số càng cao thì mức độ ảnh hưởng của kết nối này đến mạng nơ-ron càng lớn.
- Mỗi nơ-ron sẽ có một chức năng kích hoạt, về cơ bản chịu trách nhiệm "chuẩn hóa" đầu ra từ nơ-ron này.
- Dữ liệu được người dùng nhập vào mạng thần kinh đi qua tất cả các lớp và trả về kết quả ở layer cuối cùng, được gọi là lớp đầu ra (output layer).
- Trong quá trình đào tạo mô hình mạng nơ-ron, các trọng số sẽ được thay đổi và nhiệm vụ của mô hình là tìm tập giá trị của các trọng số sao cho phán đoán đúng nhất.

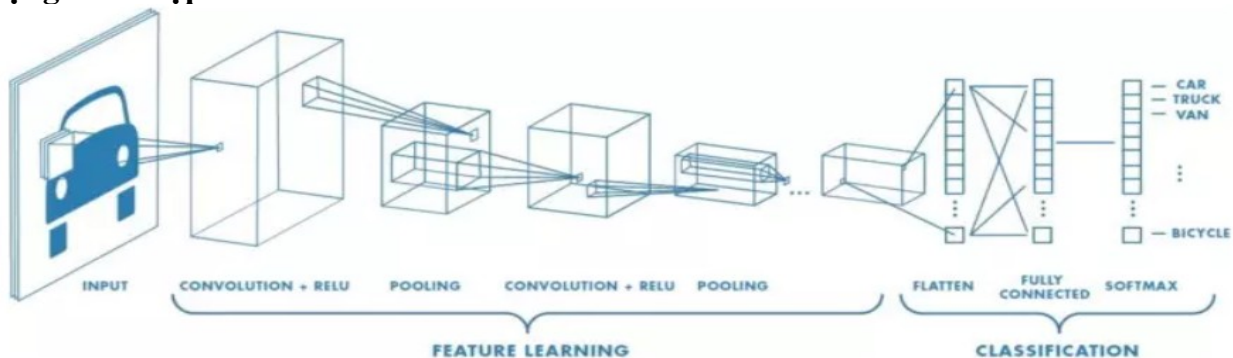
c, Computer Vision



Một dự án Computer Vision sẽ bao gồm các bước cơ bản sau đây:

- 1. Thu thập dữ liệu hình ảnh:** Thu thập một lượng lớn hình ảnh hoặc video. Các hình ảnh này có thể đã được gắn nhãn hoặc không gắn nhãn (supervised hoặc unsupervised). Gắn nhãn là việc xác định các đối tượng trong hình ảnh (ví dụ: "mèo", "xe hơi", v.v.).
- 2. Xử lý dữ liệu hình ảnh:** Dữ liệu hình ảnh sau khi thu thập sẽ được xử lý để chuẩn bị cho việc phân tích. Các thuật toán xử lý hình ảnh thực hiện các bước như điều chỉnh độ sáng, cắt xén (cropping), thay đổi kích thước và loại bỏ nhiễu.
- 3. Trích xuất đặc trưng (Feature Extraction):** Các thuật toán Computer Vision sẽ trích xuất các đặc trưng quan trọng từ hình ảnh, như các cạnh, góc, hình dạng, ... Một trong các phương pháp nổi tiếng là Convolutional Neural Networks (CNN).
- 4. Huấn luyện mô hình (Model Training):** Các mô hình học máy (VD: CNN) sẽ được huấn luyện trên tập dữ liệu hình ảnh đã qua xử lý. Mô hình sẽ học cách phân loại hoặc nhận dạng các đối tượng trong ảnh dựa trên các đặc trưng đã trích xuất.
- 5. Dự đoán và phân loại:** Dựa vào các đặc trưng đã học, mô hình sẽ đưa ra dự đoán về các đối tượng xuất hiện trong ảnh mới (chưa được gắn nhãn). Quá trình này diễn ra tự động mà không cần sự can thiệp của con người.

4. Mạng tích chập CNN

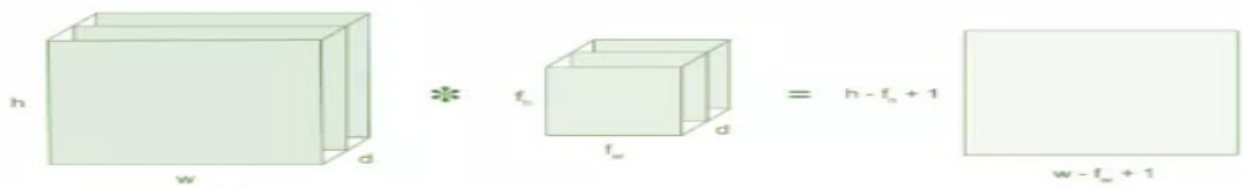


Về kỹ thuật, mô hình CNN để training và kiểm tra, mỗi hình ảnh đầu vào sẽ chuyển nó qua 1 loạt các lớp tích chập với các bộ lọc (Kernels), tổng hợp lại các lớp được kết nối đầy đủ (Full Connected) và áp dụng hàm Softmax để phân loại đối tượng có giá trị xác suất giữa 0 và 1

A, Lớp tích chập - Convolution Layer

Tích chập là lớp đầu tiên để trích xuất các tính năng từ hình ảnh đầu vào. Tích chập duy trì mối quan hệ giữa các pixel bằng cách tìm hiểu các tính năng hình ảnh bằng cách sử dụng các ô vuông nhỏ của dữ liệu đầu vào.

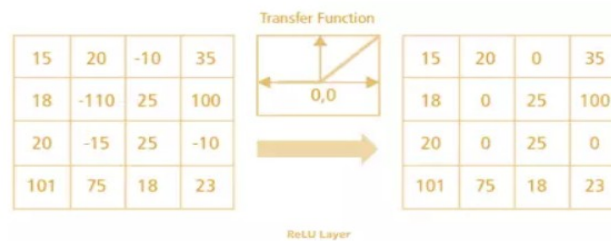
- An image matrix (volume) of dimension **(h x w x d)**
- A filter **(f_h x f_w x d)**
- Outputs a volume dimension **(h - f_h + 1) x (w - f_w + 1) x 1**



B, Hàm phi tuyến – ReLU

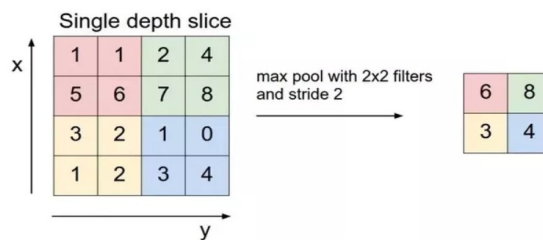
ReLU viết tắt của Rectified Linear Unit, là 1 hàm phi tuyến. Với đầu ra là: $f(x) = \max(0, x)$.

- Tại sao ReLU lại quan trọng: ReLU giới thiệu tính phi tuyến trong ConvNet. Vì dữ liệu trong thế giới mà chúng ta tìm hiểu là các giá trị tuyến tính không âm.



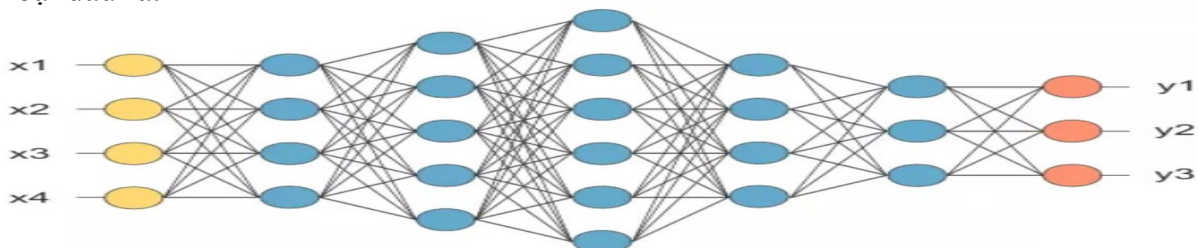
C, Lớp gộp - Pooling Layer

Lớp pooling sẽ giảm bớt số lượng tham số khi hình ảnh quá lớn. Không gian pooling còn được gọi là lấy mẫu con hoặc lấy mẫu xuống làm giảm kích thước của mỗi map nhưng vẫn giữ lại thông tin quan trọng.



D, Fully connected layer

Sau khi ảnh được truyền qua nhiều convolutional layer và pooling layer thì model đã học được tương đối các đặc điểm của ảnh thì tensor của output của layer cuối cùng sẽ được là phẳng thành vector và đưa vào một lớp được kết nối như một mạng nơ-ron. Với FC layer được kết hợp với các tính năng lại với nhau để tạo ra một mô hình. Cuối cùng sử dụng softmax hoặc sigmoid để phân loại đầu ra.



Tài liệu tham khảo:

<https://viblo.asia/p/deep-learning-tim-hieu-ve-mang-tich-chap-cnn-maGK73bOKj2>