**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**

**PHÂN HIỆU TẠI TP. HỒ CHÍ MINH**

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO THỰC TẬP CHUYÊN MÔN**

**ĐỀ TÀI: NGHIÊN CỨU VỀ THUẬT TOÁN CNN VÀ ỨNG DỤNG TRONG VIỆC NHẬN DIỆN VĂN BẢN TỪ HÌNH ẢNH**

Giảng viên hướng dẫn: TRẦN THỊ DUNG

Sinh viên thực hiện: NGUYỄN TRUNG KIÊN

Lớp : CÔNG NGHỆ THÔNG TIN

Khoá : 58

Tp. Hồ Chí Minh, tháng 8 năm 2020

**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI**

**PHÂN HIỆU TẠI TP. HỒ CHÍ MINH**

**BỘ MÔN CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO THỰC TẬP CHUYÊN MÔN**

**ĐỀ TÀI: NGHIÊN CỨU VỀ THUẬT TOÁN CNN VÀ ỨNG DỤNG TRONG VIỆC NHẬN DIỆN VĂN BẢN TỪ HÌNH ẢNH**

Giảng viên hướng dẫn: TRẦN THỊ DUNG

Sinh viên thực hiện: NGUYỄN TRUNG KIÊN

Lớp : CÔNG NGHỆ THÔNG TIN

Khoá : 58

Tp. Hồ Chí Minh, tháng 8 năm 2020

**LỜI CẢM ƠN**

**NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN (size 15, bold)**

|  |
| --- |
| ***Tp. Hồ Chí Minh, ngày ….… tháng ….… năm ….…***  **Giáo viên hướng dẫn**  **Trần Thị Dung** |

Chương 1: mở đầu

* 1. Tổng quan về xử lý ảnh và nhận dạng văn bản

Nhận dạng là lĩnh vực được các nhà khoa học rất quan tâm vì tính ứng dụng của nó trong cuộc sống hiện nay, có nhiều lĩnh vực nhận dạng như nhận dạng tín hiệu, nhận dạng giọng nói, nhận dạng hình ảnh,… Vấn đề nhận dạng văn bản từ hình ảnh là một vấn đề thú vị đối với các nhà nghiên cứu.

Văn bản từ hình ảnh có thể được ứng dụng trong các cơ quan, nhà máy, xí nghiệp, trường học. Điển hình như hệ thống nhận diện biển số xe của các nhà giữ xe, nhận diện biển số để phạt nguội khi vi phạm giao thông. Và điển hình nhất là ứng dụng trong việc dịch bằng máy ảnh của ông lớn Google.

Để nhận dạng văn bản có nhiêu phương pháp xử lý và kỹ thuật khác nhau như: logic mờ, giải thuật di truyền, mô hình xác suất thống kê, mô hình mạng nơ-ron.

Hiện nay, với sự phát triển không ngừng của khoa học máy tính, cũng như sự bùng nổ trí tuệ nhân tạo, phương pháp Deep Learning ra đời cũng dùng để phục vụ trong việc nhận dạng và xử lý ảnh. Deep Learning là thuật toán nằm trong nhóm các thuật toán Machine Learning để phát triển trí tuệ nhân tạo, Deep Learning dựa trên ý tưởng từ não bộ tới việc tiếp thu nhiều tầng biể đạt, cả cụ thể lẫn trừu tượng, qua đó làm rõ dữ liệu. Deep Learning được ứng dụng trong nhận diện hình ảnh, nhận diện giọng nói, xử lý ngôn ngữ tự nhiên. Hiện nay rất nhiều các bài toán nhận dạng sử dụng Deep Learning để giải quyết do Deep Learning có thể giải quyết các bài toán với số lượng lớn, kích thước đầu vào lớn với hiệu năng cũng như độ chính xác vượt trội so với các phương pháp phân lớp truyền thống. Các thuật toán Deep Learning phổ biến như: Deep Boltzmann Machines (DBM), Deep Belief Networks (DBN), Convolutional Neural Network (CNN) ,…

Vì những lý do trên, tôi chọn nghiên cứu đề tài: **“Nghiên cứu về thuật toán CNN và ứng dụng trong việc nhận dạng văn bản từ hình ảnh”**

* 1. Mục tiêu nghiên cứu

Mục tiêu chính của để tài này là sử dụng kỹ thuật Convolutional Neural Network (CNN) để xây dựng chương trình nhận dạng chữ viết tay

* 1. Đối tượng và phạm vi nghiên cứu
     1. Đối tượng
* Các văn bản từ hình ảnh.
* Cơ sở lý thuyết về nhận dạng ảnh.
* Các phương pháp, giải thuật về nhận dạng.
* Thuật toán Convolutional Neural Network (CNN).
  + 1. Phạm vi nghiên cứu
* Nghiên cứu kỹ thuật xử lý ảnh
  1. Ý nghĩa của đề tài

Về khoa học: Đề tài sẽ mang ý nghĩa cung cấp tài liệu về mặt lý thuyết để làm rõ về các phương pháp và kỹ thuật nhận diện văn bản từ hình ảnh.

Về thực tiễn: Đề tài góp phần hỗ trợ trong việc phát triển các ứng dụng liên quan đến việc xử lý ảnh không chỉ nằm trong nhận diện chữ viết như nhận diện đồ vật,…

* 1. Cấu trúc báo cáo thực tập chuyên môn
  + Chương 1: Mở đầu
  + Chương 2: Tìm hiểu về xử lý ảnh và nhận dạng hình ảnh.
  + Chương 3: Tổng quan về thuật toán CNN – Convolutional Neural Network.
  + Chương 4: Nhận dạng văn bản với CNN
  + Chương 5: Kết luận và hướng phát triển

Chương 2: Tìm hiểu về xử lý ảnh và nhận dạng hình ảnh

* 1. Tổng quan về nhận dạng hình ảnh
     1. Giới thiệu chung
     2. Các bước trong quá trình xử lý ảnh
        1. Thu nhận ảnh
        2. Tiền xử lý
        3. Phân đoạn
        4. Tách các đặc tính
        5. Nhận dạng ảnh
  2. Các phương pháp xử lý ảnh
     1. Nhị phân ảnh
     2. Phát hiện biên
  3. Các phương pháp nhận dạng văn bản từ hình ảnh

Có nhiều phương pháp trong các hệ thống nhận dạng chữ số viết tay, có thể kể đến như: đối sánh mẫu, thống kê, cấu trúc, mạng nơ-ron, SVM. Trong bài báo cáo này tôi chọn dùng mạng nơ-ron (thuật toán CNN trong Deep Learning)

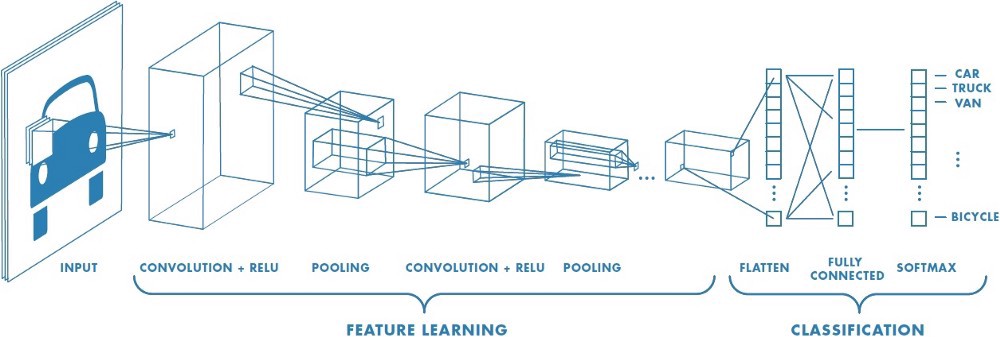
Chương 3: Tổng quan về thuật toán CNN – Convolutional Neural Network

1. Tổng quan về mạng nơ-ron tích trập

Convolutional Neural Network (CNN – Mạng nơ-ron tích chập)là một trong những mô hình Deep Learning. Nó giúp cho chung ta xây dựng được hệ thống thông minh với độ chính xác khá cao. Như hệ thống xử lý ảnh lớn như Facebook, Google, Amazon,… đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe tự lái, drone giao hàng tự động. Đáng chú ý là sự phát triển của FaceID – nhận diện mở khóa bằng khuôn mặt do các “ông lớn” trong lĩnh vực di động phát triển như Iphone, Huawei,…

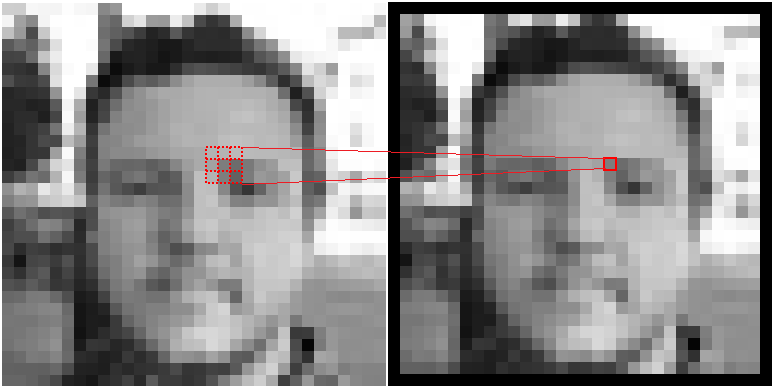
CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh. Để tìm hiểu tại sao thuật toán này được sử dụng rộng rãi, chúng ta hãy cùng tìm hiểu thuật toán này

1. Covolutional (tích chập) là gì?

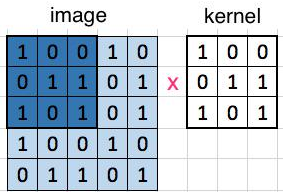
Convolution là một kỹ thuật quan trọng xử lý ảnh. Nó có mặt trong hầu hết các thuật toán làm mờ (Gausian Blur), hay làm rõ các đường (Edge detector).Có thể hiểu tích chập là một cửa sổ trượt trên một ma trận. Trong nhận dạng ảnh, convolution layer (tạm dịch là lớp tích trập) là một tầng biến đổi ma trận đầy vào để làm rõ và tách ra cá đặc tính của hình ảnh mà vẫn bảo toàn tính tương quan không gian giữa đầu ra và đầu vào

*Hình 3.1. Mạng Deep Learning gồm nhiều Convolution layer và Pooling layer*

Các convolutional layer có các tham số (kernel) đã được học để tự điều chỉnh lấy ra những thông tin chính xác nhất mà không cần chọn các đặc tính.

1. Phương pháp tích trập

*Hình 3.2. Một ví dụ cho phép tích trập*

Như hình trên tôi xin phép lấy một ví dụ dể nhìn hơn bằng ma trận 5x5 và có cửa sổ chập là một ma trận 3x3:

Hình 3.3. Ví dụ cho cửa sổ trượt

Dựa vào hình trên, ta có bên trái là ma trận đầu vào, có thể là ảnh. Bên phải là kernel là một ma trận dùng để biến đổi ma trận đầu vào trong 2 vòng lặp lồng nhau convolution. Từ trái qua phải, từ trên xuống dới dịch chuyển cửa sổ có kích thước bằng với kernel trên ma trận đầu vào, thực hiện phép nhân từ phần tử ở cùng vị trí ở ma trận cửa sổ với kernel rồi tính tổng giá trị scalar và điền vào ma trận kết quả. Cứ thế cho đến khi chạy hết:

1. Những phương pháp tích trập
2. Cấu trúc của CNN

Chương 4: Nhận dạng văn bản với CNN

1. Đặt vấn đề bài toán
2. Mô hình tổng quát
3. Xây dựng mạng nơ-ron nhận dạng ký tự
4. Xây dựng mạng nơ-ron
5. Phân tích ảnh
6. Xây dựng chương trình thử nghiệm
7. Kết quả thu được

Chương 5: Kết luận và hướng phát triển