

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

— * —

ĐỒ ÁN
TỐT NGHIỆP ĐẠI HỌC
NGÀNH CÔNG NGHỆ THÔNG TIN

**Ứng dụng học máy trong việc tự động xác
định địa điểm du lịch nổi tiếng**

Sinh viên thực hiện : **Lê Văn Mạnh**
Lớp KSVB2 – K37
Giáo viên hướng dẫn: **GV.Đinh Viết Sang**

HÀ NỘI 6-2019

Mục lục

CHƯƠNG 1 - MỞ ĐẦU	4
1.1. Nhiệm vụ của đồ án.....	4
1.2. Phương pháp thực hiện.....	4
1.3. Ý nghĩa khoa học và thực tiễn.....	4
1.4. Kết quả dự kiến	4
CHƯƠNG 2 - CONVOLUTIONAL NEURAL NETWORK	5
2.1. Giới thiệu về CNN.....	5
2.2. CNN hoạt động như thế nào.....	5
2.3. Convolution là gì	6
2.4. Stride và Padding.....	8
2.5. Pooling là gì.....	8
2.6. Fully connected neural network	9
CHƯƠNG 3 – BÀI TOÁN NHẬN DẠNG ĐỊA DANH	9
3.1. Đầu vào và đầu ra của bài toán	9
3.2. Một số kiến trúc mạng convolutional neural network	10
CHƯƠNG 4 – TRIỂN KHAI NHẬN DẠNG ĐỊA DANH.....	16
4.1. Thử nghiệm dựa trên kiến trúc mạng LeNet-5.....	17
4.2. Thử nghiệm dựa trên kiến trúc mạng AlexNet	18
4.3. Thử nghiệm dựa trên kiến trúc mạng VGG16	18
4.4. Thử nghiệm dựa trên kiến trúc mạng BNet.....	18
CHƯƠNG 6 - THIẾT KẾ HỆ THỐNG.....	19
6.1. Biểu đồ ca sử dụng	19
6.2. Biểu đồ lớp	19
6.3. Biểu đồ hoạt động.....	21
6.4. Biểu đồ tuần tự	23
6.5. Cơ sở dữ liệu địa danh.....	23
CHƯƠNG 7 – ĐÁNH GIÁ KẾT QUẢ	25
7.1. Giao diện trưng trình	25

7.2.	Minh họa chức năng tìm kiếm địa danh bằng hình ảnh	25
7.3.	Độ chính xác của hệ thống	25
7.4.	Hướng phát triển trong tương lai.....	25
CHƯƠNG 8 - TÀI LIỆU THAM KHẢO		26

CHƯƠNG 1 - MỞ ĐẦU

1.1. Nhiệm vụ của đồ án

Hiện nay để biết về một địa danh hay một điểm du lịch thông thường người dùng sẽ lên các trang tìm kiếm ví dụ google.com, bing.com ...sau đó gõ từ khóa tên hoặc địa điểm du lịch muốn tới và sau đó đọc các thông tin liên quan tới địa điểm du lịch. Tuy nhiên, với sự bùng nổ của các mạng xã hội, các ứng dụng di động cùng với sự đa dạng về loại dữ liệu đặc biệt là dữ liệu ảnh nhiều trường hợp người dùng chỉ có một bức ảnh về địa điểm du lịch hoặc muốn tới một nơi có phong cảnh đẹp như trong bức ảnh mình đang có. Điều trên dẫn tới nhu cầu tìm kiếm thông tin địa danh thông qua hình ảnh ngày càng phổ biến.

Đồ án thực hiện trên dữ liệu ảnh về các địa điểm du lịch nổi tiếng, mỗi địa danh sẽ chứa khoảng 1000 ảnh kèm với thông tin về địa lý liên quan tới địa danh đó. Do giới hạn về thời gian và nền tảng phần cứng nên hệ thống xây dựng để nhận diện và gợi ý 64 địa điểm du lịch khác nhau trên lãnh thổ Việt Nam.

1.2. Phương pháp thực hiện

Với bài toán nhận diện thông tin qua ảnh việc lập trình truyền thống sẽ khó có độ chính xác cao do độ phức tạp và đặc thù của thông tin dữ liệu. Qua một thời gian tìm hiểu công nghệ, em lựa chọn phương pháp xây dựng hệ thống với phần lõi nhận diện ảnh sẽ sử dụng trí tuệ nhân tạo để đạt độ chính xác cao nhất có thể.

1.3. Ý nghĩa khoa học và thực tiễn

Người dùng khi xem ảnh có thể ngay lập tức tìm kiếm thông tin địa danh thông qua hình ảnh mình đang xem. Mọi thứ sẽ trở lên nhanh chóng và thuận tiện cho người sử dụng góp phần thúc đẩy ngành dịch vụ và du lịch của Việt Nam.

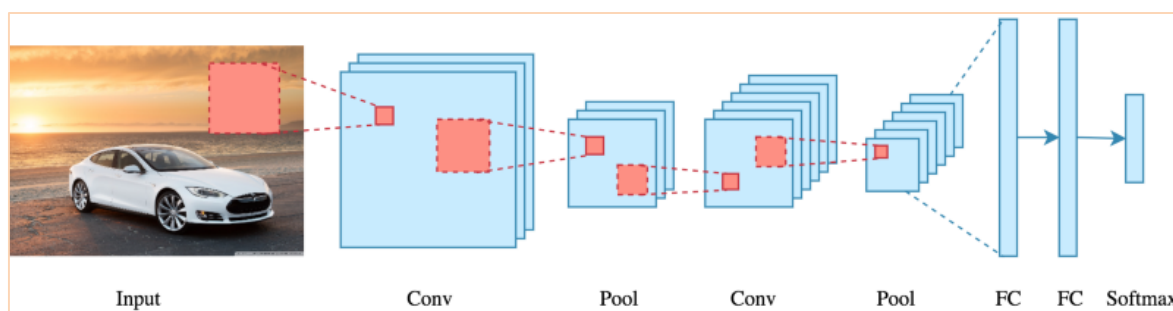
1.4. Kết quả dự kiến

Xây dựng mạng neuron nhân tạo nhận diện 64 địa danh thông qua hình ảnh với độ chính xác trên 80%, triển khai hệ thống trên nền tảng web cho người dùng truy cập và upload ảnh lên sau đó trả lại kết quả cho người dùng trên giao diện website.

CHƯƠNG 2 - CONVOLUTIONAL NEURAL NETWORK

2.1. Giới thiệu về CNN

Convolutional Neural Network (CNN – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay như hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động. CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh.



Hình 2.2.1 Sơ đồ tổng quát CNN

2.2. CNN hoạt động như thế nào

Local receptive fields: Trong mạng neural network truyền thống mỗi một neural trong input layer kết nối với một neural trong hidden layer. Tuy nhiên trong CNN chỉ một vùng xác định trong các neural trong input layer kết nối với một neural trong hidden layer. Những vùng xác định nêu trên gọi là Local receptive fields. Sự kết nối giữa input layer và hidden được chính là việc từ Local receptive fields trên một ảnh đầu vào được biến đổi thông qua một phép toán được gọi là convolution để thu được một điểm trên hidden layer.

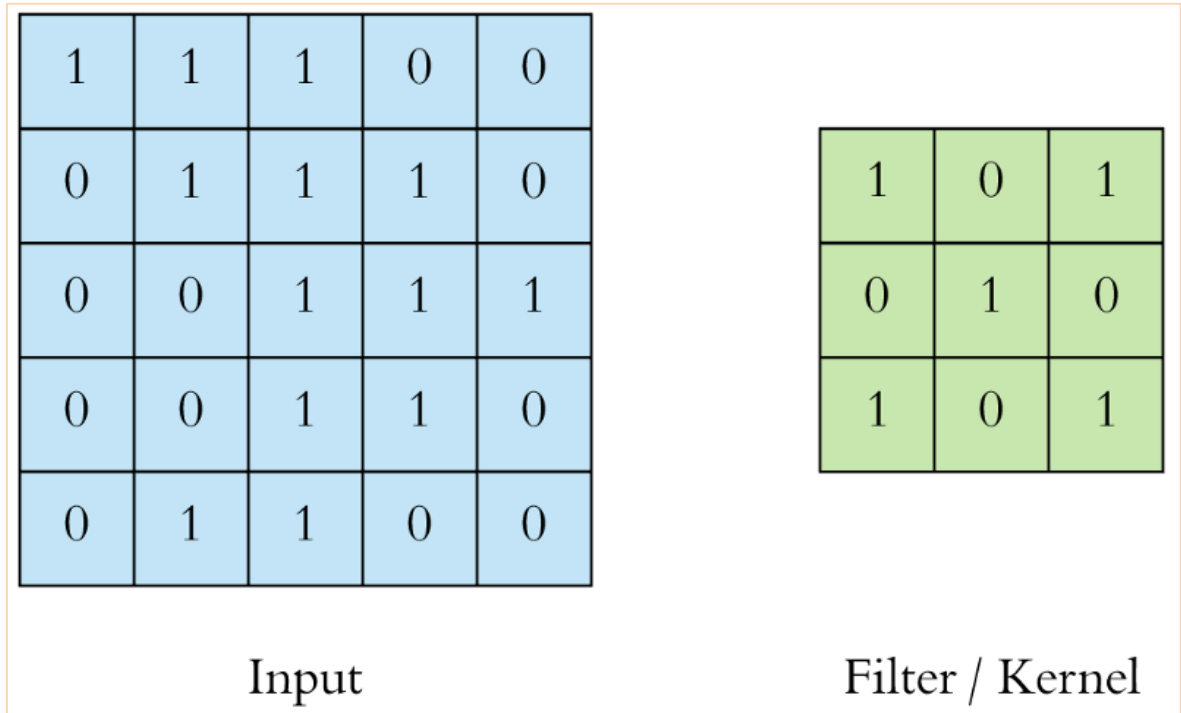
Shared weights và biases: Giống với mạng neural network truyền thống CNN cũng có tham số weights và biases. Các tham số này được học trong suốt quá trình training và liên tục cập nhật giá trị với mỗi mẫu mới (new training example). Tuy nhiên, các trọng số trong CNN là giống nhau đối với mọi neural trong cùng một lớp (layer). Điều này có nghĩa là tất cả các hidden neural trong cùng một lớp đang cùng tìm kiếm trung một đặc trưng (ví dụ như cạnh của ảnh) trong các vùng khác nhau của ảnh đầu vào.

Activation và pooling: Activation là một bước biến đổi giá trị đầu ra của mỗi neural thông qua việc sử dụng một số hàm ví dụ hàm ReLU. Giá trị thu được

sau phép biến đổi là giá trị dương nhất có thể của output, trong trường hợp output mang giá trị âm thì giá trị nhận được là 0.

pooling là một bước nhằm giảm số chiều của ma trận, các thức phổ biến nhất là từ một vùng trên ma trận ta chọn ra số có giá trị lớn nhất làm kết quả thu được sau bước pooling (max pooling)

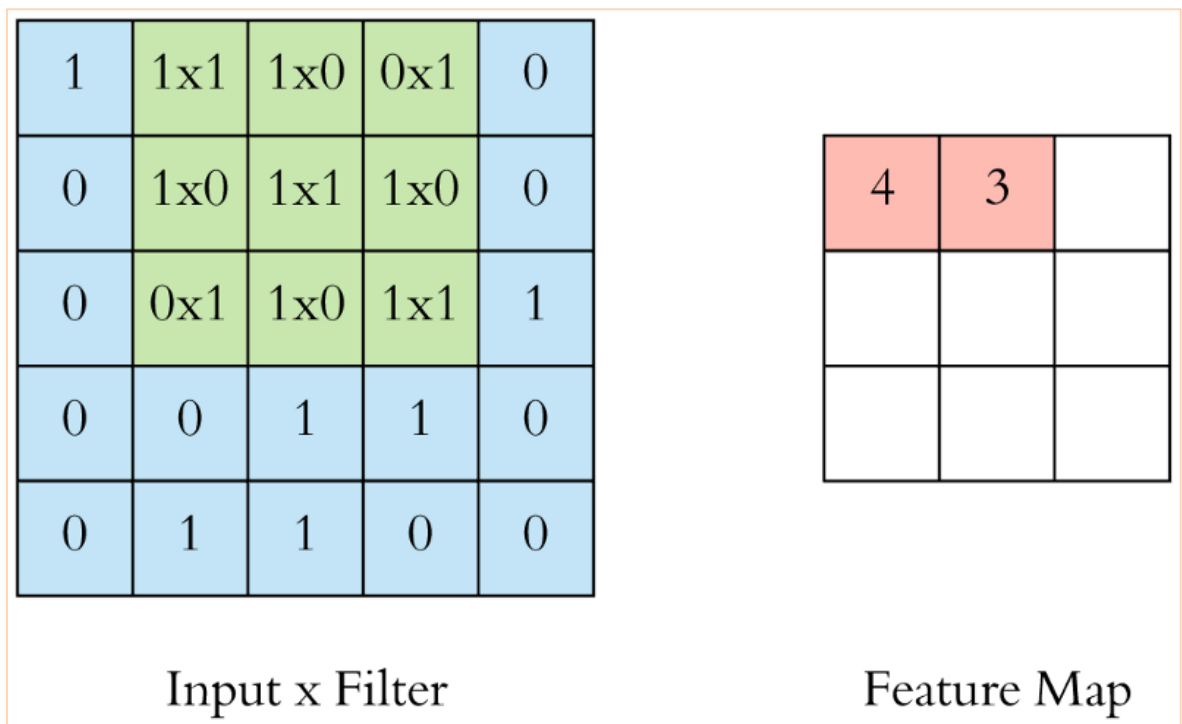
2.3. Convolution là gì



Hình 2.4.1 Convolutional là gì

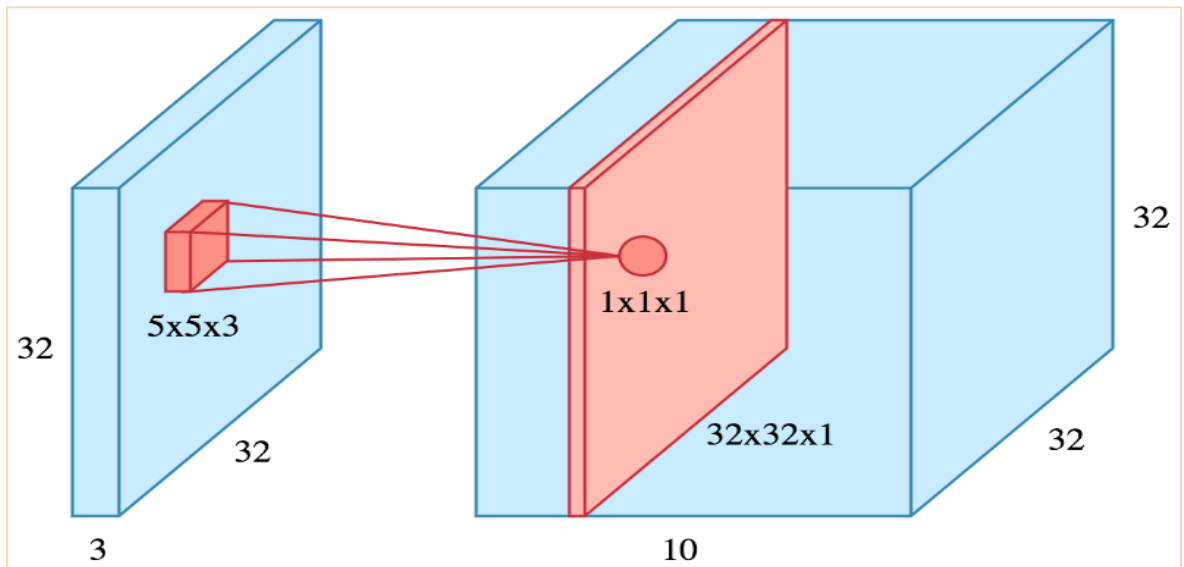
Khối cơ bản tạo lên CNN là convolutional layer. Convolution là một phép toán học để kết hợp hai khối thông tin với nhau. Trong trường hợp này, convolution được áp dụng trên dữ liệu đầu vào (ma trận) và sử dụng một mặt nạ gọi là convolution filter để tạo ra một mảng mới gọi là feature map.

Việc thực hiện phép toán convolution được mô tả như hình dưới đây với đầu vào là một mảng hai chiều 5x5 phần tử là filter có kích thước là 3x3 phần tử. Cửa sổ filter sẽ được trượt từ trái qua phải, từ trên xuống dưới. Tại mỗi vị trí của cửa sổ filter ta thực hiện nhân tương ứng từng phần tử trong ma trận đầu vào với từng phần tử trong filter, sau đó cộng tổng các tích với nhau ta thu được kết quả là một phần tử trên feature map. Quá trình thực hiện được mô tả trong hình minh họa sau đây.



Hình 2.4.2 Convolutional và mảng hai chiều

Trên đây là mô tả thực hiện phép toán convolution với ma trận hai chiều với một filter duy nhất. Trong thực tế đối với ảnh RGB ta thực hiện convolution với ma trận ba chiều ví dụ như ảnh RGB và với cùng một ảnh đầu vào ta áp dụng phép toán convolution với nhiều filter khác nhau. Mỗi một filter được áp dụng cho ta một feature layer. Nhiều feature layer xếp chồng lên nhau ta thu được một convolution layer. Ví dụ sau thể hiện ảnh có kích thước 32x32 và có ba kênh màu, ta sử dụng 10 filter và thu được convolution layer là một ma trận 32x32x10.



Hình 2.4.3 Convolutional và mảng ba chiều

2.4. Stride và Padding

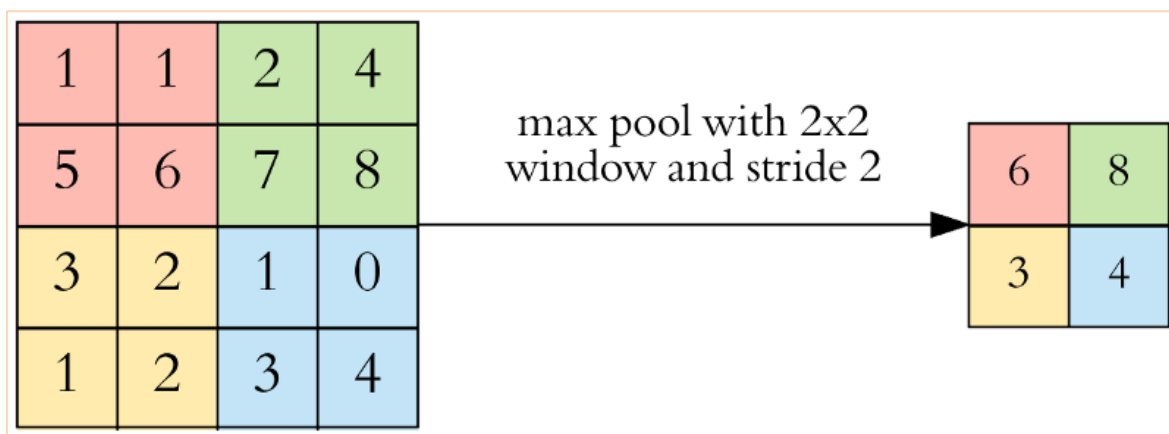
Stride là số bước nhảy của mỗi lần dịch chuyển convolution filter, trong ví dụ đầu tiên về convolution ta nhận thấy kích thước của feature map nhỏ hơn kích thước của ma trận đầu vào. Để kích thước của feature map bằng kích thước của ma trận đầu vào ta cần phải bổ xung thêm một số điểm bao quanh ma trận đầu vào thường là thêm các phần tử 0 vào xung quanh ma trận đầu vào, thao tác trên được gọi là padding.

Khi thực hiện phép toán convolution với đầu vào là ma trận vuông có kích thước là $n \times n$, stride là s , kích thước filter là $f \times f$, vùng padding có kích thước là p ta có kích thước của feature map thu được là:

$$Output\ size = \left(\frac{n + 2p - f}{s} + 1 \right) \times \left(\frac{n + 2p - f}{s} + 1 \right)$$

2.5. Pooling là gì

Sau khi thực hiện phép toán convolution chúng ta thường sử dụng pooling nhằm giảm số chiều của dữ liệu. Loại pooling thông dụng nhất là max pooling tức là trong một vùng được chọn (pooling window) được chọn của ma trận, ta lấy phần tử có kích thước lớn nhất, cũng giống với convolution thì pooling window cũng được định nghĩa kích thước (size) và bước nhảy (stride). Dưới đây là ví dụ việc áp dụng max pooling sử dụng 2×2 window và stride là 2.



Hình 2.6.1 Minh họa max pooling

2.6. Fully connected neural network

Sau các lớp convolution + pooling layers chúng ta sẽ gắn thêm một mạng Artificial Neural Network nhằm phục vụ quá trình training và nhận diện đối tượng. Thông qua quá trình học CNN tự động sẽ cập nhật lại giá trị cho các filter, weight matrix W và bias b .

CHƯƠNG 3 – BÀI TOÁN NHẬN DẠNG ĐỊA DANH

3.1. Đầu vào và đầu ra của bài toán

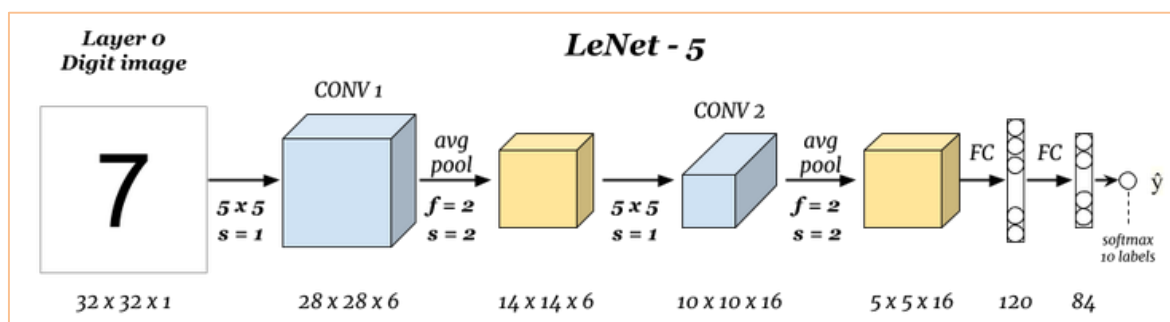
Đầu vào bài toán em chia thành hai dạng đó là đầu vào dữ liệu (dataset) dùng cho việc huấn luyện mạng neuron và đầu vào của bài toán cần phải nhận diện.

Dữ liệu đầu vào dùng cho huấn luyện mạng hay còn gọi là dataset là tập ảnh có kích thước 480x480x3, đây là ảnh RGB về các địa danh du lịch nổi tiếng của Việt Nam. Như đã nói ở phần mở đầu trong phần phạm vi của đề án, dữ liệu đầu vào chứa 64 loại ảnh về 64 địa danh tương ứng khác nhau. Các ảnh được lưu trữ trong từng thư mục khác nhau với mỗi thư mục chứa khoảng 1000 ảnh. Quá trình huấn luyện mạng sẽ chia tập dữ liệu thành hai phần, trong đó 70% dùng cho việc huấn luyện mạng và 30% dùng cho việc kiểm định độ chính xác của mạng. Trong quá trình huấn luyện và kiểm định độ chính xác của ảnh, các ảnh sẽ được thay đổi kích thước một cách hợp lý để phù hợp với thiết kế của mạng neuron. Do tập ảnh với số lượng 1000 cho mỗi địa danh là chưa đủ lớn để tăng số lượng ảnh cho việc huấn luyện em sẽ sử dụng một số kỹ thuật như lật, xoay ảnh gốc để thu được một ảnh khác nhằm tăng số lượng ảnh cho việc huấn luyện.

Dữ liệu đầu vào dành cho việc nhận dạng địa danh là ảnh của địa danh có kích thước bất kỳ với định dạng ảnh yêu cầu là RGB.

3.2. Một số kiến trúc mạng convolutional neural network

LeNet-5 là một mạng cổ điển và cơ bản nhất cho bài toán nhận diện ảnh, được ứng dụng cho bài toán nhận diện số và chữ viết tay với đặc điểm dữ liệu đầu là ảnh có độ chi tiết đơn giản và kích thước nhỏ. Kiến trúc được công bố trong bài báo của Y. Lecun, L. Bottou, Y. Bengio and P. Haffner vào năm 1998.



Hình 3.2.1 Sơ đồ kiến trúc mạng LeNet-5

Sau đây là bảng tóm tắt kiến trúc của mạng, gồm các lớp cùng với các tham số cần phải học tương ứng.

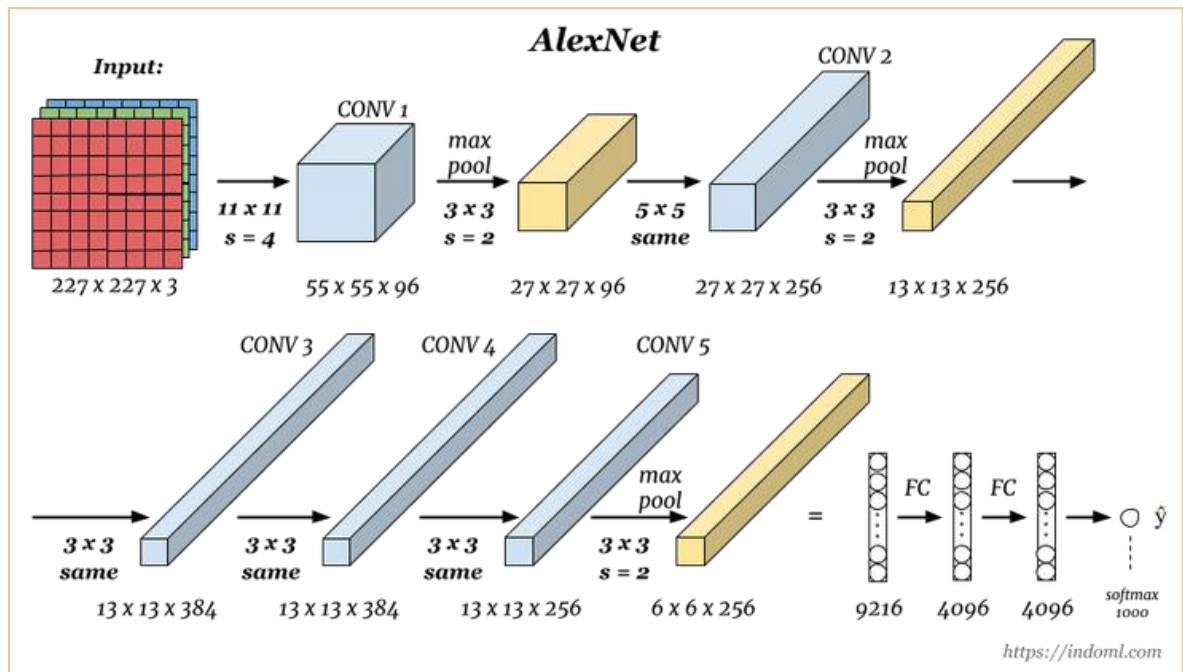
Layer	Filters	Biases	Stride	Padding	Tensor	Parameters
Input	0	0	0	0	$32 \times 32 \times 1$	0
Conv-1	$5 \times 5 \times 1 \times 6$	6	1	0	$28 \times 28 \times 6$	156
MaxPool-1	2×2	0	2	0	$14 \times 14 \times 6$	0
Conv-2	$5 \times 5 \times 6 \times 16$	16	1	0	$10 \times 10 \times 16$	2416
MaxPool-2	2×2	0	2	0	$5 \times 5 \times 16$	0
FC-1	400×120	120	0	0	120	48120
FC-2	120×84	84	0	0	84	10164
RBF	0	0	0	0	10	0
Total						60856

Bảng 3.2.1 Kiến trúc mạng LeNet-5

Đầu vào là ảnh Gray với kích thước 32×32 pixel và đầu ra là khoảng các Euclidean giữa mỗi vector đầu vào vector trọng số tức tensor đầu ra của FC-2.

Ta có thể thấy đây là mạng rất đơn giản với kích thước đầu vào cũng như số lượng các tham số phải học. Tiếp theo chúng ta tìm hiểu một kiến trúc phổ biến khác nhưng phức tạp hơn mạng LeNet-5.

Mạng AlexNet. Dùng để phân loại 1000 loại ảnh có kích thước $227 \times 227 \times 3$. Kiến trúc được công bố bởi Alex Krizhevsky, Geoffrey Hinton, and Ilya Sutskever vào năm 2012



Hình 3.2.2 Sơ đồ kiến trúc mạng AlexNet

Sau đây là bảng tóm tắt các thông số của mạng chứa thông tin về kích thước tensor đầu ra của từng lớp cùng với số lượng tương ứng các tham số mà mạng cần phải học.

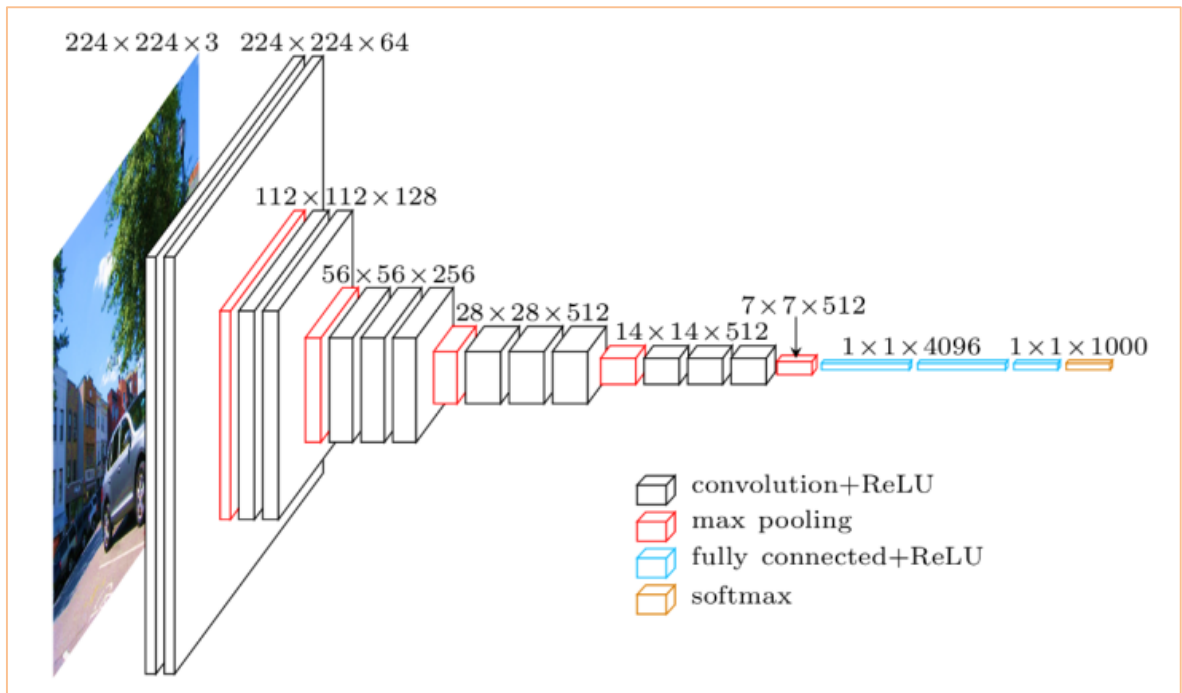
Layer	Filters	Biases	Stride	Padding	Tensor	Parameters
Input	0	0	0	0	$227 \times 227 \times 3$	0
Conv-1	$11 \times 11 \times 3 \times 96$	96	4	0	$55 \times 55 \times 96$	34944
MaxPool-1	3×3	0	2	0	$27 \times 27 \times 96$	0
Norm-1	0	0	0	0	$27 \times 27 \times 97$	0

Conv-2	5x5x96x256	256	1	2	27x27x256	614656
MaxPool-2	3x3	0	2	0	13x13x256	0
Norm-2	0	0	0	0	13x13x256	0
Conv-3	3x3x256x384	384	1	1	13x13x384	885120
Conv-4	3x3x384x384	384	1	1	13x13x384	1327488
Conv-5	3x3x384x256	256	1	1	13x13x256	884992
MaxPool-3	3x3	0	2	0	6x6x256	0
FC-1	9216x4096	4096	0	0	4096	37752832
FC-2	4096x4096	4096	0	0	4096	16781312
FC-3	4096x1000	1000	0	0	1000	4097000
Output	0	0	0	0	1000	
Total						62378344

Bảng 3.2.2 kiến trúc mạng AlexNet

Mạng với đầu vào là ảnh có kích thước 227x227x3 và kết quả đầu cần thực hiện là phân loại 100 ảnh khác nhau, tổng số tham số cần phải học của mạng là 62378344 lớn hơn so với mạng LeNet-5. Các trọng số được cập nhập thông qua quá trình training mạng bởi thuật toán backpropagation.

Mạng VGG-16 được công bố trong bài báo của Karen Simonyan and Andrew Zisserman vào năm 2014.



Hình 3.2.3 Sơ đồ kiến trúc mạng VGG16

Thông tin các tham số về mạng được mô tả tron bản sau, với đầu vào của mạng là ảnh RGB và đầu ra là vector chứa 1000 phần tử tương ứng với 1000 class được phân loại.

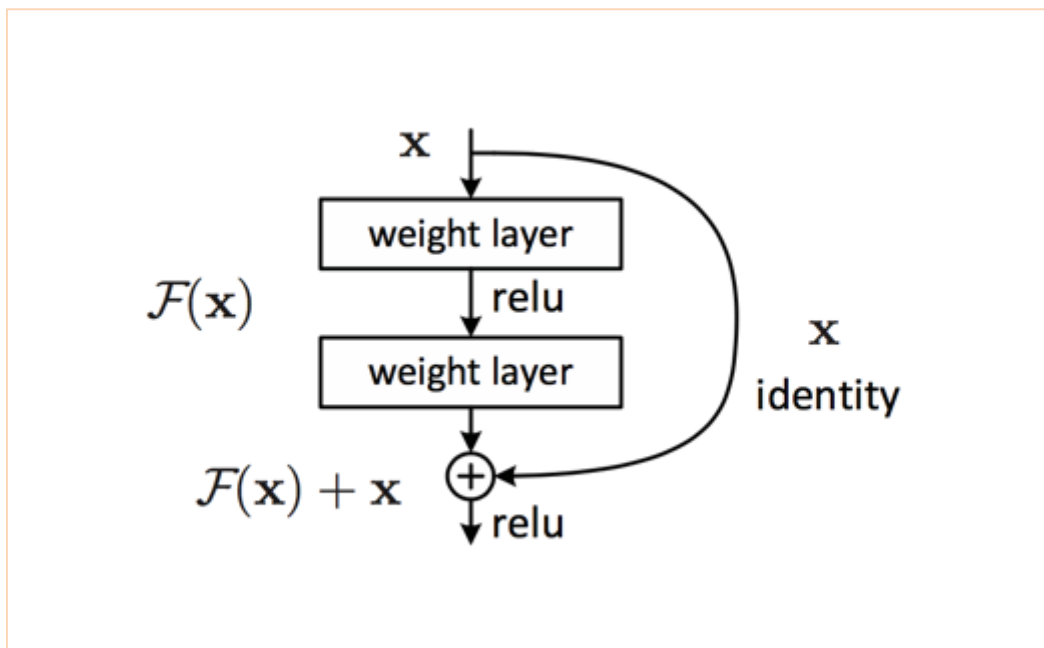
Layer	Filters	Biases	Stride	Padding	Tensor	Parameters
Input	0	0	0	0	$224 \times 224 \times 3$	0
Conv-11	$3 \times 3 \times 64$	64	1	1	$224 \times 224 \times 64$	1792
Conv-12	$3 \times 3 \times 64 \times 64$	64	1	1	$224 \times 224 \times 64$	36928
MaxPool-1	2×2	0	2	0	$112 \times 112 \times 64$	0
Conv-21	$3 \times 3 \times 64 \times 128$	128	1	1	$112 \times 112 \times 128$	73856
Conv-22	$3 \times 3 \times 128 \times 128$	128	1	1	$112 \times 112 \times 128$	147584
MaxPool-2	2×2	0	2	0	$56 \times 56 \times 128$	0
Conv-31	$3 \times 3 \times 128 \times 256$	256	1	1	$56 \times 56 \times 256$	295168
Conv-32	$3 \times 3 \times 256 \times 256$	256	1	1	$56 \times 56 \times 256$	590080
Conv-33	$3 \times 3 \times 256 \times 256$	256	1	1	$56 \times 56 \times 256$	590080

MaxPool -3	2x2	0	2	0	28x28x256	0
Conv-41	3x3x256x51 2	512	1	1	28x28x512	1180160
Conv-42	3x3x512x51 2	512	1	1	28x28x512	2359296
Conv-43	3x3x512x51 2	512	1	1	28x28x512	2359296
MaxPool -4	2x2	0	2	0	14x14x512	0
Conv-51	3x3x512x51 2	512	1	1	14x14x512	2359296
Conv-52	3x3x512x51 2	512	1	1	14x14x512	2359296
Conv-53	3x3x512x51 2	512	1	1	14x14x512	2359296
MaxPool -5	2x2	0	2	0	7x7x512	0
FC-1	25088x4096	4096	0	0	4096	102764544
FC-2	4096x4096	4096	0	0	4096	16781312
FC-3	4096x1000	1000	0	0	1000	4097000
Soft- max					1000	0
Total						138354984

Bảng 3.2.3 kiến trúc mạng VGG16

Ta có thể thấy số lượng các tham số phải học trong mạng VGG16 với cùng một đầu ra là 1000 class đã lớn hơn mạng AlexNet 2 lần.

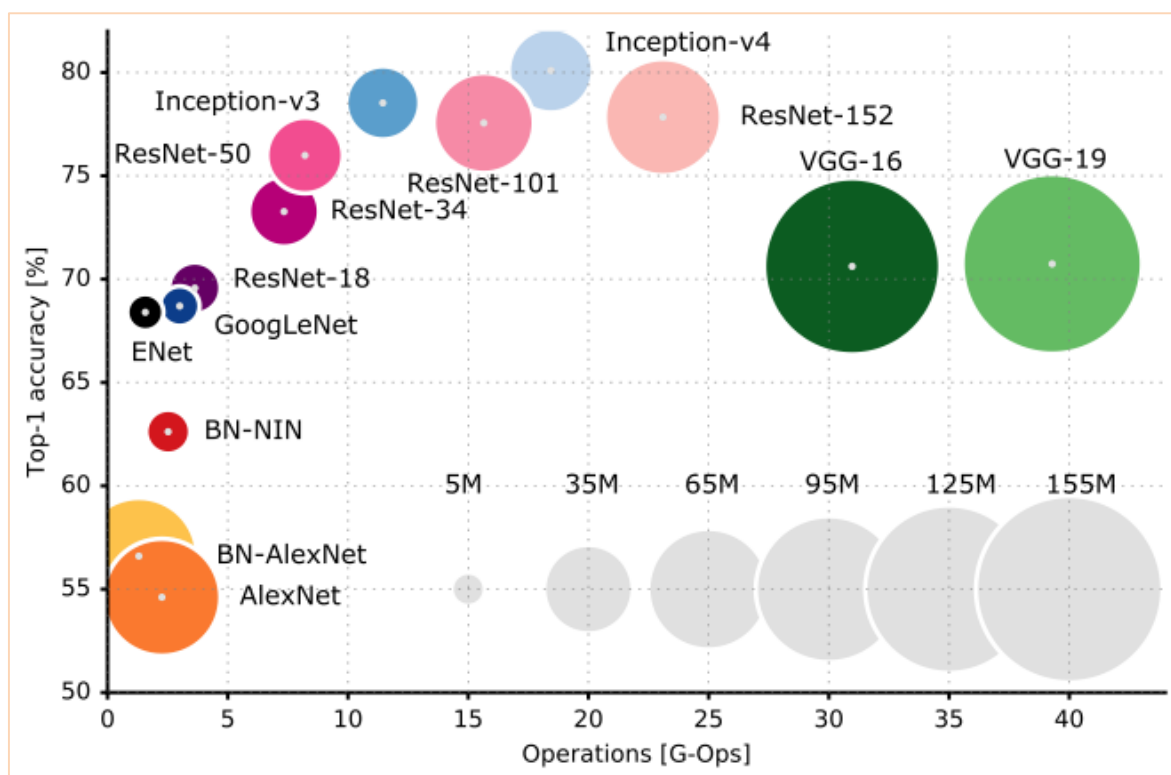
Mạng ResNet được công bố bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun vào năm 2015. Đặc điểm của mạng là số lượng layer lớn, với phần tử cơ bản có tên gọi là Residual Block được mô tả trong hình dưới đây.



Hình 3.2.4 Sơ đồ phân tử Residual Block



Hình 3.2.5 Sơ đồ kiến trúc mạng ResNet



Hình 3.2.6 So sánh độ chính xác và khối lượng tính toán

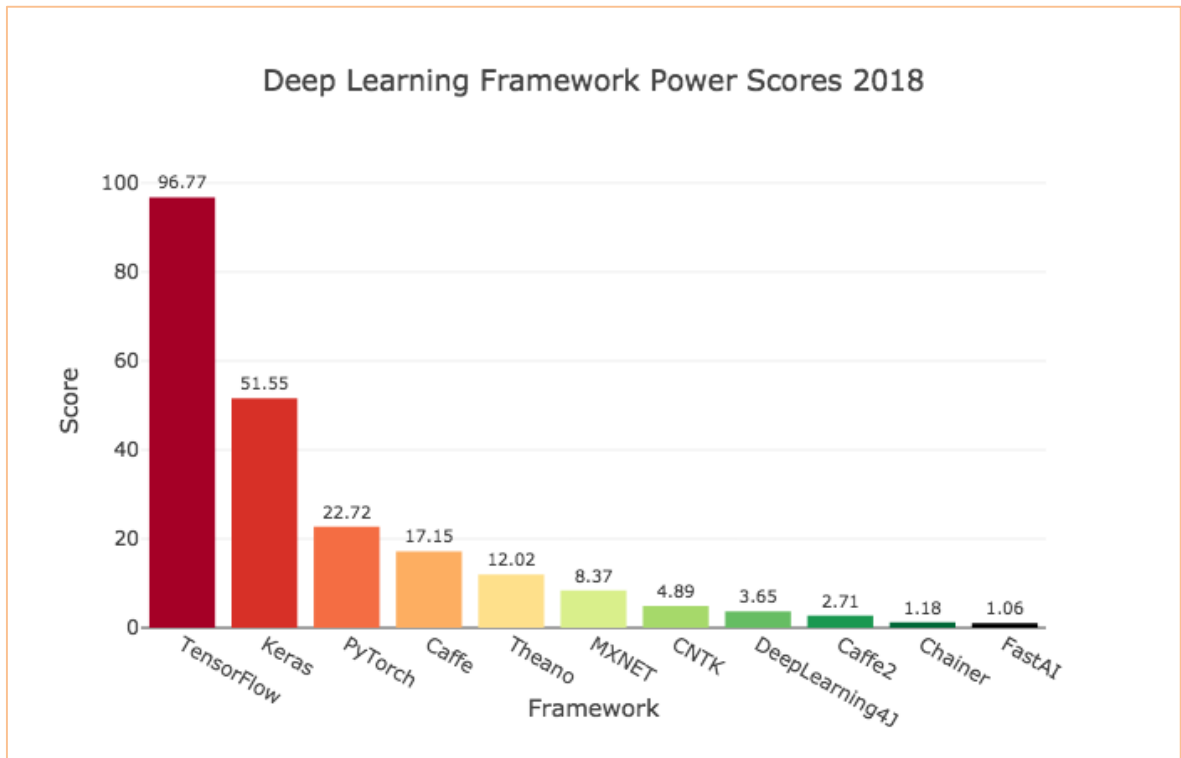
Hình trên được lấy từ kaggle.com, chúng ta có thể thấy được sự so sánh về khối lượng tính toán và độ chính xác giữa các cách xây dựng mạng CNN phổ biến hiện nay, trong đó ResNet 50 có độ chính xác khá cao trong khi khối lượng tính toán không nhiều. AlexNet cần khối lượng tính toán thấp nhưng độ chính xác tương đối thấp so với các kiến trúc mạng khác.

Trên đây là một số kiến trúc mạng phổ biến trên thế giới đã được xây dựng và kiểm nghiệm cũng như dùng trong các cuộc thi nhận diện ảnh lớn. Dựa vào các kiến trúc đã có cũng như cơ sở lý thuyết về machine learning em đi tiến hành xây dựng và kiểm nghiệm với mạng được tùy biến đối với bài toán cụ thể của em là nhận dạng địa danh.

CHƯƠNG 4 – TRIỂN KHAI NHẬN DẠNG ĐỊA DANH

Việc xây dựng và triển khai một mạng CNN trên máy tính là một công việc phức tạp đối với lập trình viên, ngoài ra mạng phải đảm bảo tận dụng tối đa nền tảng phần cứng của máy tính đặc biệt là GPU của máy tính do khối lượng tính toán trong quá trình huấn luyện mạng rất lớn. Vì lý do trên để đạt được mục tiêu xây dựng được mạng CNN trong khoảng thời gian giới hạn của việc thực hiện đồ án này em đã tìm hiểu một số thư viện hỗ trợ triển khai việc xây

dụng các thuật toán học máy. Trong số đó có một số thư viện rất phổ biến được đưa ra trong hình sau.



Hình 4.0.6 các framework được dùng trong deep learning

Hình tên được lấy trên trang towardsdatascience.com, trong bài “Deep Learning Framework Power Scores 2018” cùng với nhiều nguồn khác có thể thấy TensorFlow là một thư viện được dùng rất phổ biến. TensorFlow là một thư viện rất nổi tiếng được phát triển bởi Google được sử dụng để xây dựng lên các hệ thống học máy, thư viện này được sử dụng phổ biến trong công nghiệp phù hợp cho việc xây dựng lên các hệ thống lớn hoặc hệ thống mà lập trình viên muốn có sự tùy biến cao.

Từ các đặt điểm trên em lựa chọn TensorFlow cho việc xây dựng mạng CNN của mình.

Do giới hạn của phần cứng máy tính nên em lựa chọn đầu vào của mạng có kích thước là ảnh 256 x 256 pixel.

4.1. Thử nghiệm dựa trên kiến trúc mạng LeNet-5

Bảng thiết kế chi tiết các lớp của mạng với bài toán nhận diện địa danh.

Bảng 4.1.1 Thiết kế model dựa trên LeNet-5

Mạng được thiết kế với đầu vào là ảnh có kích thước 128x128 với 3 kênh màu RGB. Và đầu ra của mạng là một vector 64 phần tử, mỗi phần tử là giá trị sắc xuất ảnh đầu vào rơi vào địa danh có mã tương ứng.

Sau đây là việc triển khai, training và thực hiện kiểm định độ chính xác của mạng bằng thư viện Tensorflow.

4.2. Thử nghiệm dựa trên kiến trúc mạng AlexNet

Dưới đây là bảng thiết kế các lớp của bài toán khi triển khai dựa trên mạng AlexNet.

Kiến trúc mạng được triển khai trong chương trình dùng thư viện tensorflow.

Mạng sau khi triển khai sẽ được huấn luyện với tập dữ liệu ảnh về các địa danh.

Độ chính xác của mạng được mô tả trong hình dưới đây.

4.3. Thử nghiệm dựa trên kiến trúc mạng VGG16

Dưới đây là bảng thiết kế các lớp của bài toán khi triển khai dựa trên mạng AlexNet.

Kiến trúc mạng được triển khai trong chương trình dùng thư viện tensorflow.

Mạng sau khi triển khai sẽ được huấn luyện với tập dữ liệu ảnh về các địa danh.

Độ chính xác của mạng được mô tả trong hình dưới đây.

4.4. Thử nghiệm dựa trên kiến trúc mạng BKNet

Dưới đây là bảng thiết kế các lớp của bài toán khi triển khai dựa trên mạng AlexNet.

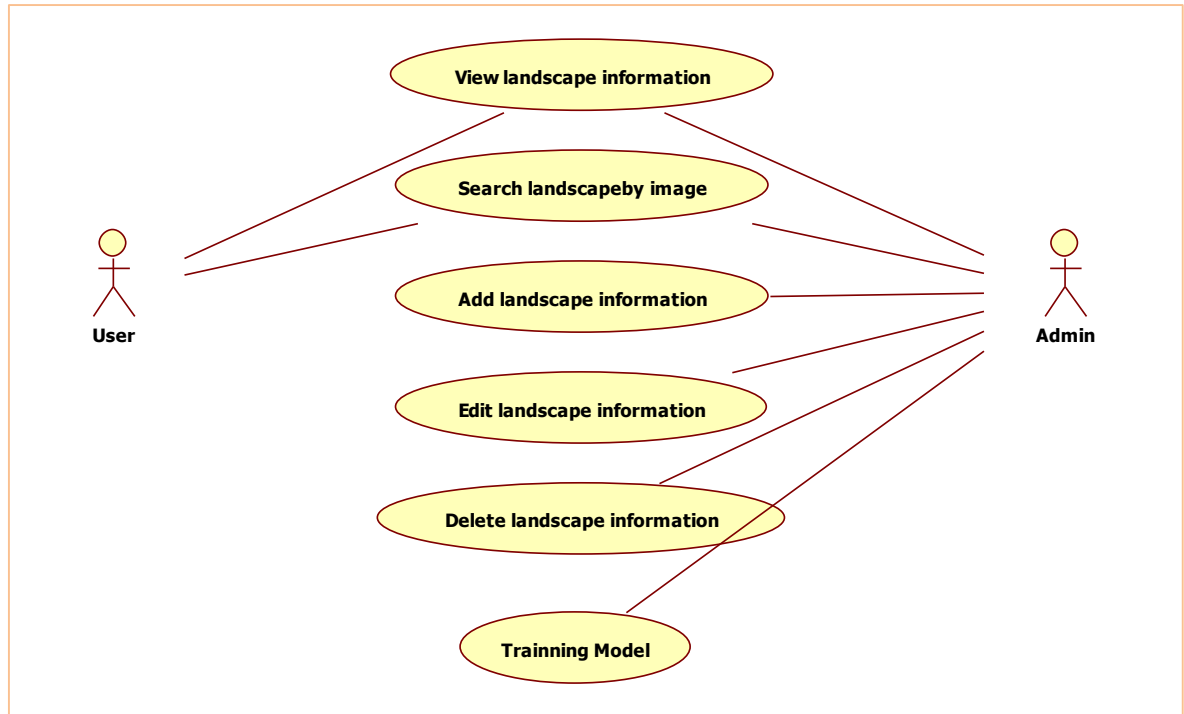
Kiến trúc mạng được triển khai trong chương trình dùng thư viện tensorflow.

Mạng sau khi triển khai sẽ được huấn luyện với tập dữ liệu ảnh về các địa danh.

Độ chính xác của mạng được mô tả trong hình dưới đây.

CHƯƠNG 6 - THIẾT KẾ HỆ THỐNG

6.1. Biểu đồ ca sử dụng



Hình 6.1.1 biểu đồ ca sử dụng

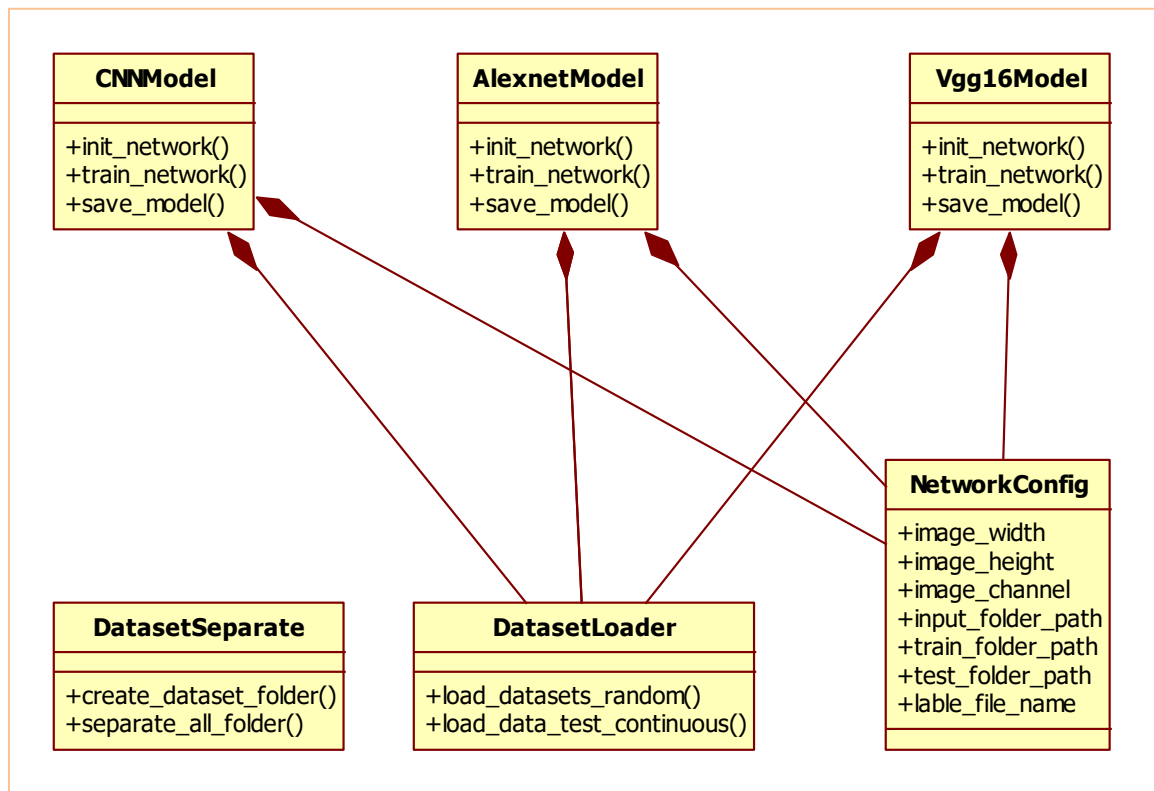
Hệ thống gồm các chức năng cơ bản giúp người dùng có thể xem thông tin và tìm kiếm địa danh bằng hình ảnh. Người quản trị có thêm các quyền liên quan tới trang quản trị như thêm, sửa, xóa thông tin địa danh.

6.2. Biểu đồ lớp

Biểu đồ lớp trong trường hợp huấn luyện mạng nhận diện địa danh gồm ba thành phần chính:

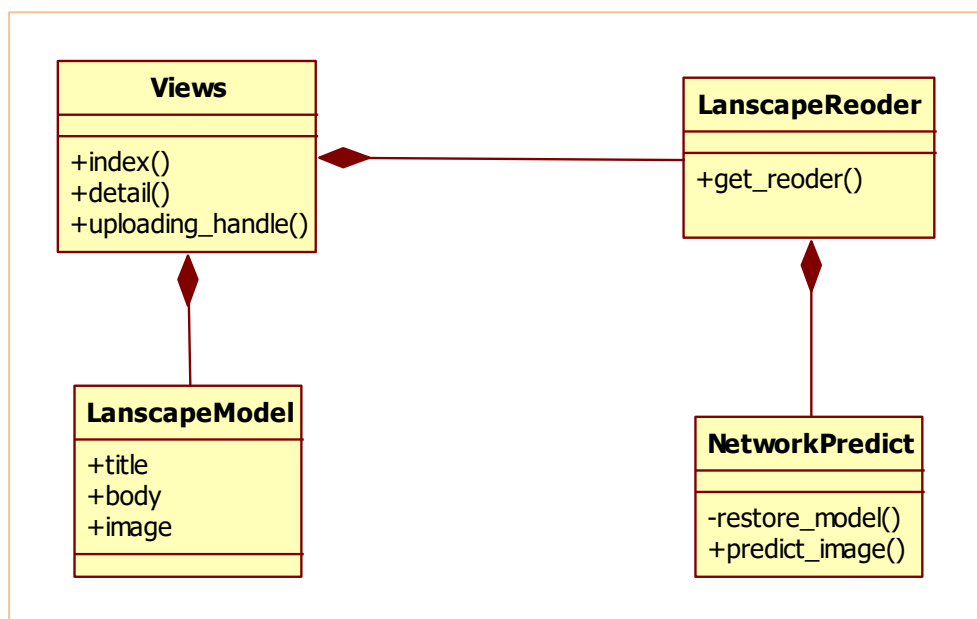
- Các class liên quan tới tiền xử lý dữ liệu như phân chia dữ liệu train và dữ liệu test, kiểm duyệt dữ liệu và loại bỏ những ảnh không đúng format ví dụ ảnh không phải rgb.
- Class liên quan tới các tham số cấu hình của mạng và tải dữ liệu ảnh từ bộ nhớ

- Class định nghĩa các model dựa trên các kiến trúc khác nhau có hàm huấn luyện và lưu lại model sau khi được huấn luyện



Hình 6.2.1 sơ đồ lớp cho user case huấn luyện mạng

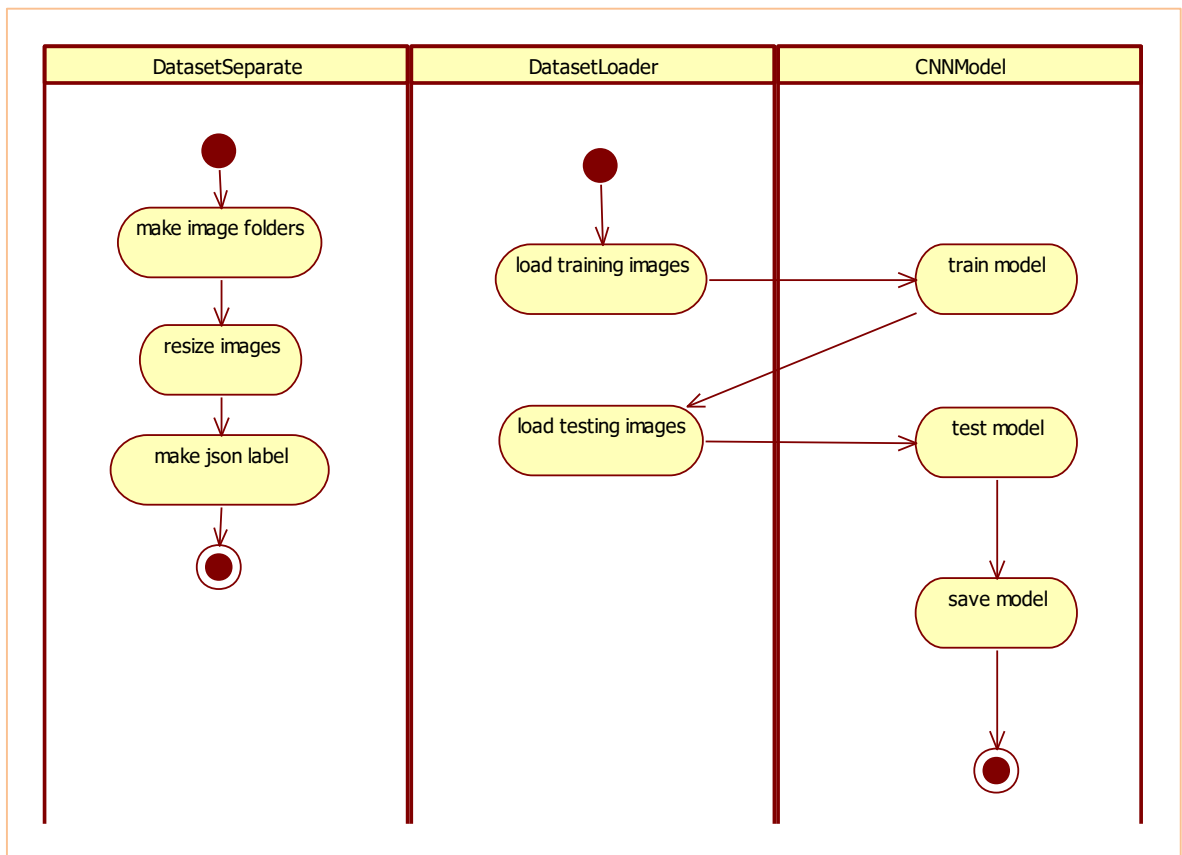
Trong trường hợp tìm kiếm thông tin bằng hình ảnh, chức năng này được thực hiện trên môi trường web. Trong đó có các lớp chính tham gia vào user case được mô tả trong hình dưới đây



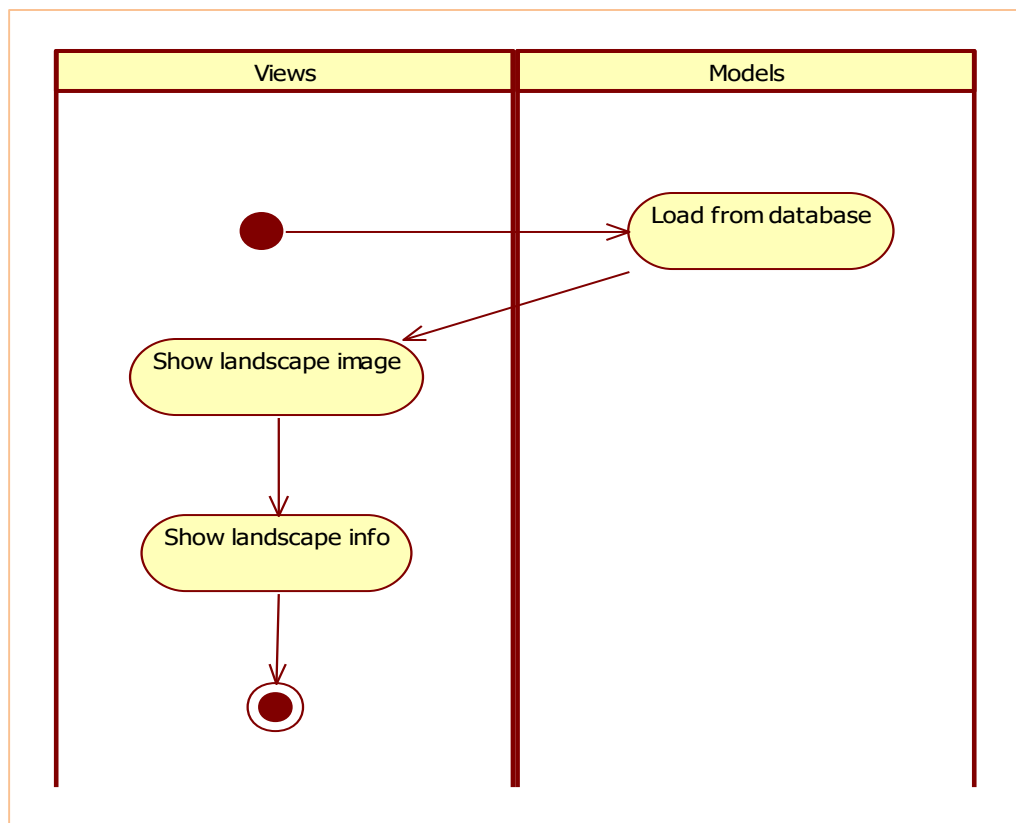
Hình 6.2.2 sơ đồ lớp cho user case tìm kiếm địa danh bằng hình ảnh

6.3. Biểu đồ hoạt động

Biểu đồ hoạt động cho việc huấn luyện mạng neuron, kết quả thu được sau quá trình huấn luyện là một model tương ứng với cấu trúc mạng đã cấu hình có thể là AlexNet hoặc VGG16Net ...

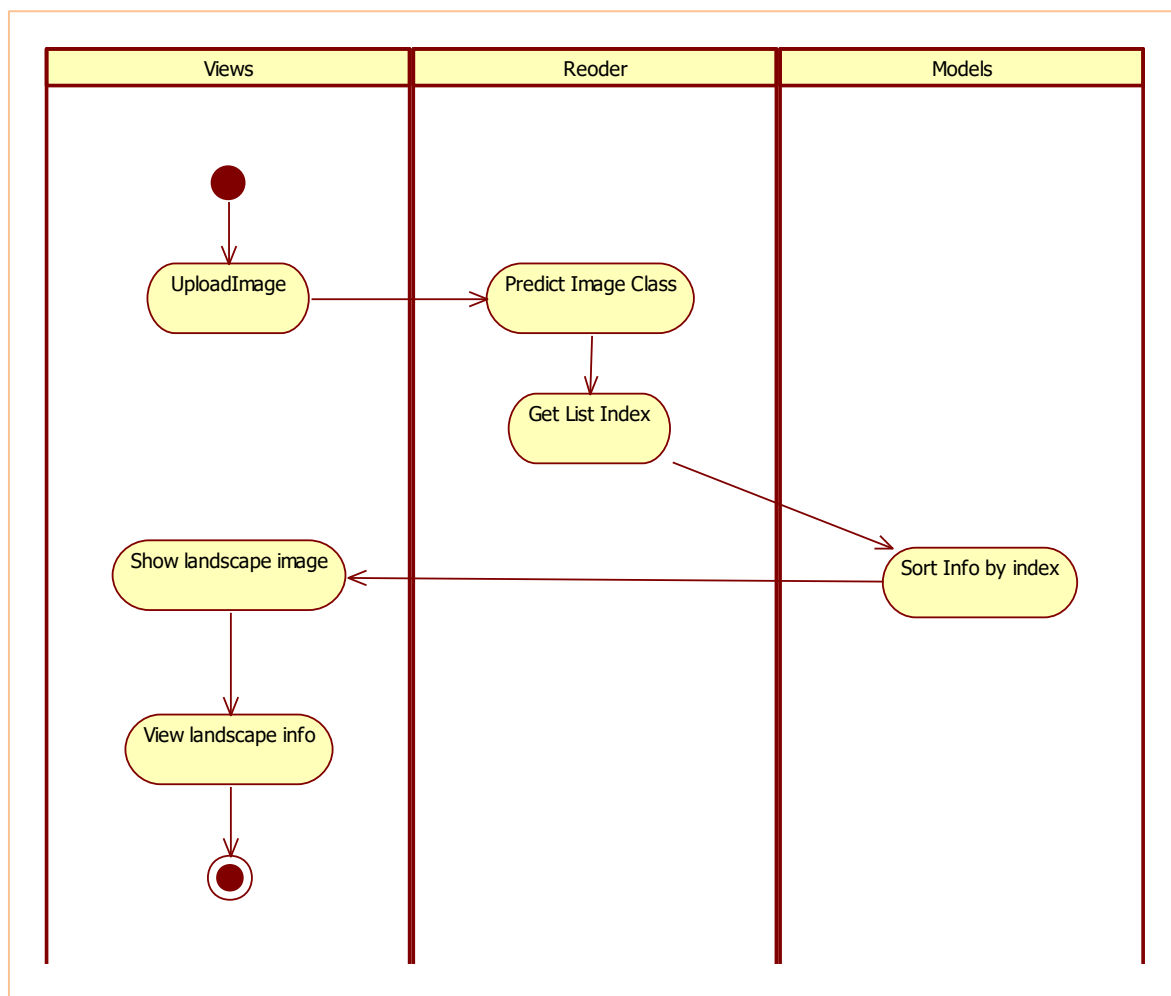


Dưới đây là biểu đồ hoạt động cho hai chức năng quan trọng nhất của hệ thống đó là chức năng xem thông tin địa danh và tìm kiếm thông tin địa danh.



Hình 6.2.1 biểu đồ hoạt động xem thông tin địa danh

Chức năng tìm kiếm thông tin địa danh bằng hình ảnh là chức năng quan trọng nhất. Trong đó sử dụng mô hình học máy được đào tạo ở phần 4 để trích chọn đặc trưng, từ đó đưa ra danh sách các địa danh có đặc điểm giống với địa danh được tìm kiếm nhất. Dựa vào đó hệ thống sẽ sắp xếp lại thứ tự danh sách các địa danh và đưa các địa danh có sắc xuất giống nhất lên vị trí đầu tiên.



Hình 6.2.2 biểu đồ hoạt động tìm kiếm thông tin địa danh

Các chức năng về thêm, sửa, xóa thông tin địa danh được Django hỗ trợ trong trang quản trị của frame work. Người quản trị đăng nhập vào hệ thống và sửa thông tin theo yêu cầu tương ứng.

6.4. Biểu đồ tuần tự

6.5. Cơ sở dữ liệu địa danh

Cơ sở dữ liệu được sử dụng trong trương trình là cơ sở dữ liệu sqlite được tích hợp sẵn trong Django Frame Work. Với phạm vi chương trình thì cơ sở dữ liệu sqlite đủ đáp ứng được nhu cầu sử dụng cũng như tính gọn nhẹ của cơ sở dữ liệu.

Dưới đây là bảng sử liệu thông tin địa danh trong cơ sở dữ liệu.

Tên trường	Kiểu dữ liệu	Ghi chú
id	integer	primary key
title	varchar(1000)	
body	text	
date	datetime	
num	varchar(3)	
image	varchar(100)	

Bảng 6.5.1 bảng dữ liệu địa danh

Trường dữ liệu “title” được dùng để lưu tên của địa danh, trường dữ liệu “body” được dùng để lưu mô tả cũng như cung cấp thông tin chi tiết về địa danh. Trường dữ liệu num để lưu thông tin về ID của địa danh, trường này dùng để tra cứu thông tin địa danh sau khi hàm nhận diện địa danh trả về kết quả là các ID của các địa danh theo thứ tự từ gần giống nhất với đặc điểm địa danh được tìm kiếm. “Image” là trường lưu đường dẫn ảnh minh họa cho địa danh, ảnh minh họa được lưu tại một thư mục trên server.

Tiếp theo là bảng chứa tài khoản người dùng của website

Tên trường	Kiểu dữ liệu	Ghi chú
id	integer	primary key
password	varchar(128)	
last_login	datetime	
is_superuser	bool	
username	varchar(150)	
first_name	varchar(30)	
email	varchar(254)	
is_staff	bool	
is_active	bool	
date_joined	datetime	
last_name	varchar(150)	

Bảng 6.5.2 bảng thông tin người dùng

Bảng được Django hỗ trợ tạo ra trong hệ thống admin, trong đó các trường quan trọng như is_superuser, is_staff được dùng để phân biệt tài khoản admin và user thông thường. Trong đề án này thông tin user em đang đề chỉ gồm một tài khoản admin để đăng nhập vào trang quản trị từ đó thêm, sửa xóa các thông tin về địa danh

CHƯƠNG 7 – ĐÁNH GIÁ KẾT QUẢ

7.1. Giao diện trưng trình

7.2. Minh họa chức năng tìm kiếm địa danh bằng hình ảnh

7.3. Độ chính xác của hệ thống

7.4. Hướng phát triển trong tương lai

CHƯƠNG 8 - TÀI LIỆU THAM KHẢO

- [1] <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>
- [2] <https://www.mathworks.com/videos/introduction-to-deep-learning-what-are-convolutional-neural-networks--1489512765771.html>
- [3] <https://medium.com/machine-learning-bites/deeplearning-series-convolutional-neural-networks-a9c2f2ee1524>
- [4] <https://www.jefkine.com/general/2016/09/05/backpropagation-in-convolutional-neural-networks/>
- [5] <http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/>
- [6] <https://www.kaggle.com/shivamb/cnn-architectures-vgg-resnet-inception-tl>
- [7] <https://www.learnopencv.com/number-of-parameters-and-tensor-sizes-in-convolutional-neural-network/>
- [8] <https://indoml.com/2018/03/07/student-notes-convolutional-neural-networks-cnn-introduction/>
- [9] <https://towardsdatascience.com/deep-learning-framework-power-scores-2018-23607ddf297a>