

**ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC BÁCH KHOA**  
**KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH**

—o0o—



**Đề cương luận văn tốt nghiệp**

## **Phát hiện kẹt xe từ camera hành trình**

**GVHD:** TS.Dương Ngọc Hiếu

Nhóm sinh viên thực hiện:

Trương Ngọc Anh 1410141

Lê Nguyễn Minh Trí 1414207

# Mục lục

<b>1</b>	<b>Giới thiệu</b>	<b>7</b>
1.1	Tổng quan . . . . .	7
1.1.1	Tính khả quan . . . . .	9
1.1.2	Ưu điểm . . . . .	9
1.1.3	Nhược điểm . . . . .	9
<b>2</b>	<b>Mạng neural và Mạng neural tích chập (Convolutional Neural Networks CNN</b>	<b>11</b>
2.1	Mạng neural (Neural Network) . . . . .	11
2.1.1	Mạng neural ( Neural Network ) . . . . .	11
2.1.2	Cấu trúc mạng neural . . . . .	12
2.1.3	Mô hình Feedforward Neural Network . . . . .	16
2.2	Mạng neural tích chập (Convolutional neural network) . . . . .	20
2.2.1	Giới thiệu mạng neural tích chập . . . . .	20
2.2.2	Mô hình mạng neural tích chập . . . . .	20
2.2.3	Kiến trúc mạng GoogLeNet) . . . . .	28
<b>3</b>	<b>Apache Hadoop</b>	<b>31</b>

3.1	Tổng quan về hệ thống Apache Hadoop . . . . .	31
3.2	Hadoop distributed file system - HDFS . . . . .	32
3.2.1	Thiết kế . . . . .	32
3.2.2	Ý tưởng chủ đạo . . . . .	32
<b>4</b>	<b>Mô hình loại ảnh giao thông</b>	<b>33</b>
4.1	Vấn đề và mục tiêu . . . . .	33
4.2	Phương pháp transfer learning . . . . .	33
4.3	Hiện thực . . . . .	34
4.3.1	Tổng quan về tập dữ liệu . . . . .	34
4.3.2	Môi trường . . . . .	35
4.4	Các bước hiện thực . . . . .	37
4.4.1	Xử lý dữ liệu . . . . .	37
4.4.2	Tạo các bottlenecks . . . . .	37
4.4.3	Huấn luyện . . . . .	38
<b>5</b>	<b>Kết quả huấn luyện và đánh giá</b>	<b>39</b>
5.1	Kết quả huấn luyện . . . . .	39
5.2	Sử dụng mô hình phân loại ảnh mới . . . . .	39
<b>6</b>	<b>Kết luận</b>	<b>44</b>
6.1	Các kết quả . . . . .	44
6.2	Hướng phát triển . . . . .	45

# Danh sách hình vẽ

2.1	Tế bào thần kinh neuron sinh học . . . . .	12
2.2	Cấu trúc cơ bản mạng neuron . . . . .	13
2.3	Đồ thị hàm step . . . . .	15
2.4	Đồ thị hàm sigmoid . . . . .	15
2.5	Đồ thị hàm Tanh . . . . .	16
2.6	Ví dụ kiến trúc mạng feedforward . . . . .	17
2.7	MLP . . . . .	18
2.8	Convolutional layer . . . . .	22
2.9	Convolutional layer . . . . .	22
2.10	Mảng các giá trị của bộ lọc . . . . .	23
2.11	ảnh đầu vào . . . . .	24
2.12	phép toán tích chập . . . . .	24
2.13	phép toán tích chập . . . . .	25
2.14	kết quả tầng tích chập . . . . .	25
2.15	max-pooling 2 x 2 . . . . .	26
2.16	Ví dụ tầng tổng hợp . . . . .	27
2.17	GoogLeNet . . . . .	28
2.18	inception module . . . . .	29

2.19	naive inception module . . . . .	29
5.1	Kết quả huấn luyện . . . . .	40
5.2	Ảnh kiểm thử 1 - kết quả . . . . .	41
5.3	Ảnh kiểm thử 2 - kết quả . . . . .	42
5.4	Ảnh kiểm thử 3 - kết quả . . . . .	43
5.5	Ảnh kiểm thử 4 - kết quả . . . . .	43

# Danh sách bảng

4.1	Các chiều dữ liệu. . . . .	35
4.2	Một vài kiểu dữ liệu. . . . .	36

# Chương 1

## Giới thiệu

### 1.1 Tổng quan

Từ những năm gần đây, tình trạng ùn tắc giao thông ở Tp.Hồ Chí Minh đã không dừng lại ở diễn biến phức tạp mà lại còn gia tăng hơn trước. Cụ thể, trên các tuyến đường hiện nay, tình trạng kẹt xe không chỉ xảy ra ở giờ cao điểm mà còn ở các khung giờ khác. Nguyên nhân dẫn đến sự việc trên, một phần ảnh hưởng bởi điều kiện thời tiết, một phần do các công trình cải tạo hạ tầng, khi thi công lấn chiếm mặt đường. Nhưng phần lớn là do mật độ xe cộ ngày một đông dần, dẫn đến việc ùn ứ, ùn tắc, di chuyển chậm,... Từ những nguyên nhân đó, nhóm sinh viên chúng em đã xây dựng mô hình hệ thống phát hiện kẹt xe từ camera hành trình (cụ thể là camera hành trình xe buýt), với mong muốn có thể góp phần giải quyết được một phần nhỏ tình trạng giao thông hiện nay ở Tp. Hồ Chí Minh nói riêng cũng như ở Việt Nam nói chung.

Thông qua từng giai đoạn xây dựng hệ thống, đã giúp cho chúng em có thể

tiếp cận, học hỏi thêm nhiều kiến thức về các công nghệ nổi tiếng hiện nay như: Deep Learning, Tensorflow, Apache Hadoop, Apache Hive,... Từ những ý tưởng ban đầu, cơ bản chỉ là việc **Phân tích một bức ảnh giao thông và nêu lên kết quả là: "Kẹt xe", "Thông thoáng"** thông qua việc huấn luyện dữ liệu trên kiến trúc *GoogleNet* - sẽ được trình bày ở chương 2, tại em đã phát triển, mở rộng ra bằng việc ứng dụng dữ liệu lớn (Big Data) như Apache Hadoop để tiến hành lưu trữ dữ liệu (bao gồm: các file video định dạng **.avi** được lấy từ camera hành trình xe buýt, các file hình ảnh được định dạng **.jpg** được cắt ra từ video, đem vào phân tích,... và một file SQL định dạng **???** được lưu trữ bằng Apache Hive với các thuộc tính như: Kinh độ, Vĩ độ, Thời gian,... , (nội dung của *Hadoop & Hive* sẽ được trình bày ở chương 3).

Trong quá trình hoàn thành từng giai đoạn, tại em đã thấy được những mặt khó khăn trong quá trình thực hiện, đó là việc vận dụng mô hình để nhận biết hình ảnh **t nghĩ phần này phải do m viết, t viết nó lủng củng lắm**.

Những dạng bài toán về việc phân tích ảnh đã có rất nhiều thuật toán, mô hình cũng như là các công nghệ đã được đề xuất và áp dụng rộng rãi. Nhóm chúng em cũng sử dụng lại một mô hình có sẵn đó chính là *GoogleNet* để giúp cho việc huấn luyện dữ liệu được tối ưu hơn (cụ thể là khi chúng ta cần mở rộng nhãn - "label" của đối tượng). Điểm khác biệt của nhóm em, đó là ngoài việc ứng dụng lại mô hình kiến trúc *GoogleNet*, tại em sẽ tích hợp vào trong Hadoop, thiết kế thành một hệ thống vừa có thể lưu trữ video, vừa có thể phân tích hình ảnh và lấy dữ liệu gps (kinh độ, vĩ độ, thời gian,...) để xuất ra thông tin về tình trạng giao thông ở vị trí đó như thế nào, đó cũng



là chính là mục đích để tụi em tiến hành và phát triển bài toán.

Bên cạnh đó, có những vấn đề khó khăn vẫn chưa giải quyết được trong bài toán, sẽ được trình bày ở phần tiếp theo.

### **1.1.1 Tính khả quan**

Để đánh giá về tính khả quan trong đề tài này, chúng ta cần phân tích một số đặc điểm. Thứ nhất là về tập ảnh giao thông, ở đây, tập ảnh giao thông tốt hay không tùy thuộc vào nhiều yếu tố như nguồn gốc tập ảnh hay cơ cấu hạ tầng camera giao thông được đầu tư như thế nào. Đối với thành phố Hồ Chí Minh nói riêng, chúng ta vẫn còn đang xây dựng hệ thống cơ sở hạ tầng giám sát nên chưa thể dựa vào nguồn camera giám sát để lấy dữ liệu. Vì thế, chúng em sẽ sử dụng tập ảnh từ camera hành trình trên các xe buýt để thực hiện xây dựng mô hình. Khi đã có tập dữ liệu đủ tốt, chúng ta có thể sử dụng mô hình được xây dựng trên đề tài này và có thể được áp dụng vào hệ thống thực tiễn.

### **1.1.2 Ưu điểm**

- Hệ thống file được lưu trữ bằng cách dùng công nghệ Apache Hadoop.

### **1.1.3 Nhược điểm**

- Ảnh lấy từ Camera hành trình còn có nhiều hạn chế.
- Dòng Stream lấy từ Video chưa thực hiện một cách triệt để.

- Mô hình hệ thống còn gặp nhiều khó khăn trong việc áp dụng vào bài toán thực tế.

## Chương 2

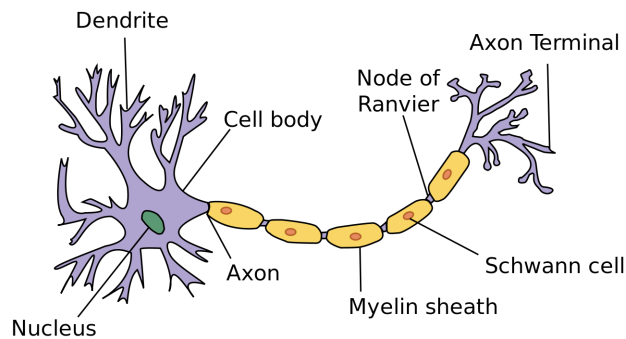
# Mạng neural và Mạng neural tích chập (Convolutional Neural Networks CNN

### 2.1 Mạng neural (Neural Network)

#### 2.1.1 Mạng neural ( Neural Network )

Theo khái niệm về sinh học, mạng neural là sự kết nối giữa các tế bào thần kinh neural lại với nhau. Trong lĩnh vực trí tuệ nhân tạo, mạng neural còn được gọi là Artificial Neural Network (ANN) - mạng neural nhân tạo, đây là mô hình xử lý dữ liệu, mô phỏng lại chức năng và cách hoạt động của hệ thống neural sinh học ở con người. Hình 2.1 minh họa cấu trúc của tế bào thần kinh neuron.

Mạng neural có gồm nhiều đơn vị kết nối, làm việc như một thể thống



Hình 2.1: Tế bào thần kinh neuron sinh học

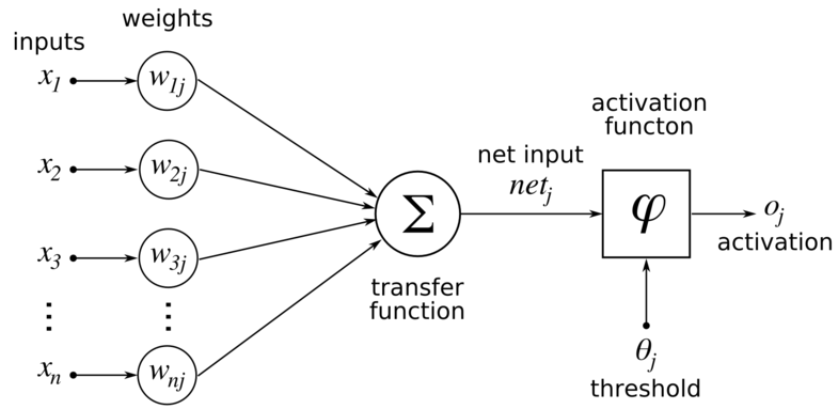
nhất thông qua việc trao đổi thông tin nhờ các liên kết.

### 2.1.2 Cấu trúc mạng neural

Như đã trình bày, các neuron trong một mạng làm việc như một thể thống nhất bằng việc trao đổi thông tin. Thực tế, đây là quá trình điều chỉnh các trọng số được truyền từ input ban đầu kết hợp với các hàm tính toán để có được các thông số trọng số phù hợp nhất. Quá trình này còn được gọi là quá trình học hay huấn luyện. Hình 2.2 mô tả cấu trúc đơn giản nhất của một mạng neuron.

#### Cấu trúc mạng neuron

- **Tập các node:** bao gồm nhiều node, mỗi node là đơn vị nhỏ nhất giữ chức năng xử lý thông tin của mạng.
- **Các tầng:** Các node trên được xếp thành các tầng, các node chung một tầng không thể kết nối nhau. Trong đó tầng input và tầng output là 2 tầng thiết yếu. Tùy vào một số mạng cụ thể có thể có thêm một hay nhiều tầng nằm ở giữa được gọi là tầng ẩn (hidden layer).



Hình 2.2: Cấu trúc cơ bản mạng neuron

- **Tầng input - input layer:** các nối ở tầng này nhận dữ liệu đầu vào và truyền tới các node ở các tầng kế tiếp, trong một số trường hợp còn có chức năng xử lý thông tin.
- **Tầng ẩn - hidden layer:** một số mạng có thể có thêm tầng ẩn, số lượng tầng ẩn trong một mạng có thể nhiều hơn 1. Có chức năng nhận các giá trị từ từng input hoặc tầng ẩn trước nó, tính toán các giá trị và gửi đến các node ở các tầng ẩn hoặc tầng output tiếp theo đó tùy theo từng mạng cụ thể.
- **Tầng output - output layer:** nhận giá trị từ tầng trước đó (tầng ẩn hoặc tầng input) để tính toán các giá trị ngõ ra.
- **Các liên kết:** mỗi node trong một tầng truyền thông tin qua các node ở các tầng khác thông qua các liên kết. Các giá trị mà các liên kết này được gán sẽ được gọi là trọng số liên kết (weight). Giá trị trọng số được kết nối vào neuron j với neuron k là  $w_{kj}$ .
- **Hàm truyền - transfer function:** dùng để tính tổng các tích input

với trọng số liên kết của nó.

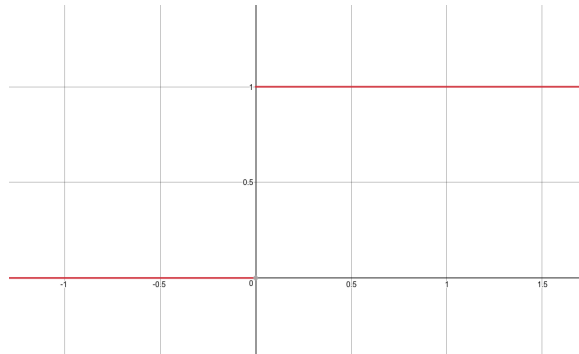
$$\sum input_j * w_{ij}$$

- **Activation function:** dùng để tính toán giá trị input sang giá trị output. Tùy vào mục đích và cụ thể từng loại mạng mà có nhiều loại activation function khác nhau.

Trong các bài toán khác nhau, người có những loại hàm activation như sau.

– *Step function:*

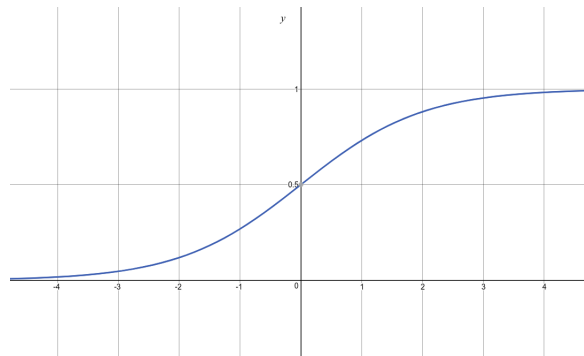
$$\begin{cases} 0 & x < 0 \\ 1 & x > 0 \end{cases}$$



Hình 2.3: Đồ thị hàm step

– *sigmoid function:*

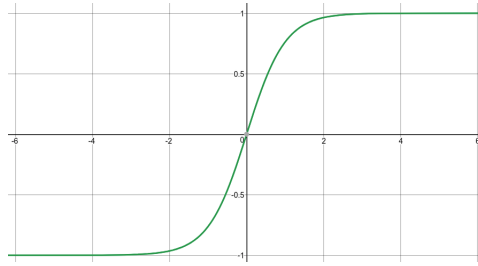
$$A(x) = \frac{1}{1 + e^{-x}}$$



Hình 2.4: Đồ thị hàm sigmoid

– *Tanh function:*

$$\text{Tanh}(x) = \frac{2}{1 + e^{-2x}} - 1$$



Hình 2.5: Đồ thị hàm Tanh

### 2.1.3 Mô hình Feedforward Neural Network

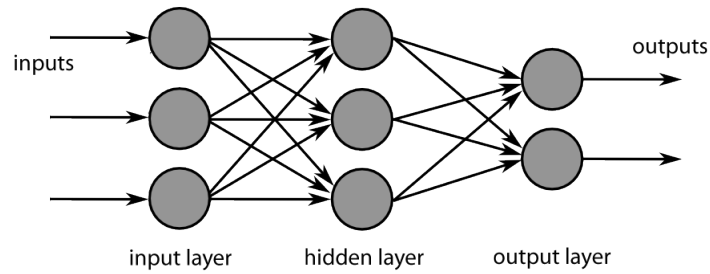
Thời điểm hiện nay, chúng ta có rất nhiều loại mô hình mạng neuron do sự khác nhau về sự kết hợp cũng như về mặt kiến trúc và thuật toán mà mạng đó áp dụng. Trong phần này, chúng ta sẽ tìm hiểu về mô hình mạng Feedforward Neural Network (FFNN), đây là kiến trúc mạng neuron được sử dụng phổ biến trong các bài toán dự báo. Mô hình gồm hai thành phần chính đó là kiến trúc feedforward - mạng truyền thẳng và giải thuật Backpropagation được áp dụng trong mạng.

#### Kiến trúc Feedforward

Đối với mạng feedforward, cấu trúc gồm một tầng input, một tầng output và có thể có nhiều hơn một tầng ẩn nằm giữa hai tầng input và output.

Như hình 2.6, một mạng Feedforward, trong tầng input và output thì số lượng neuron tại mỗi hai tầng này sẽ là cố định tùy theo đặc tính của dữ liệu. Đối với tầng ẩn, số lượng tầng ẩn cũng như số lượng neuron trong mỗi tầng tùy thuộc vào cá nhân thiết kế.





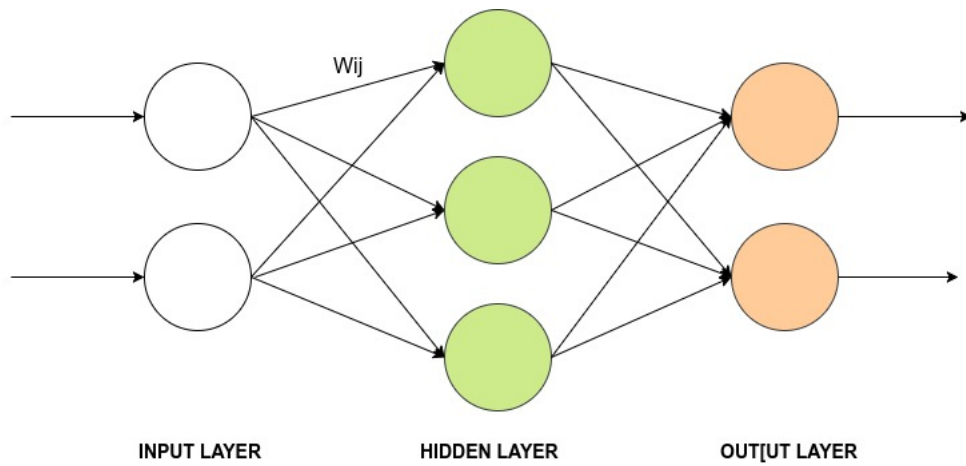
Hình 2.6: Ví dụ kiến trúc mạng feedforward

### Giải thuật Backpropagation

Ở nội dung này, chúng ta sẽ không đi sâu vào kỹ thuật xử lý các phép tính[2] đạo hàm mà chỉ trình bày giải thuật lan truyền ngược một cách đơn giản nhất.

Các ký hiệu và hàm được dùng trong trình bày giải thuật:

- $W_{ij}$ : là trọng số nối node thứ  $i$  tới node  $j$  ở layer kế tiếp.
- $I_j$ : là đầu vào tại node thứ  $j$ .
- $O_j$ : là kết quả xuất tại node thứ  $j$ .
- $\theta_j$ : bias tại node  $j$ .
- $l$ : tốc độ học của mạng (learning rate).
- $Err_j$ : giá trị lỗi tại node thứ  $j$ .
- Activation function được dùng trong nội dung này là hàm Sigmoid như mục trên



Hình 2.7: MLP

Nội dung thuật toán.

**Input:**

- Mạng feed forward với  $n$  input,  $m$  node ở tầng ẩn, và  $p$  output.
- Hệ số học hay tốc độ học  $l$ .
- Tập dữ liệu huấn luyện  $D$ .
- Sai số học  $\epsilon$ .

**Output:** Vector các trọng số mới sau khi huấn luyện.

**Nội dung thuật toán:**

- **Bước 1:** Khởi tạo ngẫu nhiên các giá trị trọng số  $W_{ij}$ .
- **Bước 2:** Tính toán các giá trị đầu vào  $I_j$  và đầu ra  $O_j$ .
  - Tại node  $i$  ở tầng input:

$$I_i = x_i, O_i = I_i$$

- Tại node  $j$  ở tầng khác:

$$I_j = \sum_i W_{ij} O_i + \theta_j$$

$$O_j = \frac{1}{1 + e^{-I_j}}$$

- **Bước 3:** Tính toán lỗi trung bình và đánh giá.

- Tại node thuộc tầng output:

$$Err_j = O_j(1 - O_j)(T_j - O_j)$$

- Tại node thuộc tầng ẩn:

$$Err_j = O_j(1 - O_j) \sum_k Err_k W_{jk}$$

Với  $Err_k, W_{jk}$  là giá trị lỗi tại node  $k$  ở tầng tiếp theo và giá trị trọng số của node  $j$  đến  $k$ .

Thuật toán sẽ dừng lại khi  $Err_k \leq \epsilon$

- **Bước 4:** Cập nhật các trọng số và độ lệch

$$W_{ij} = W_{ij} + (l)Err_j O_i$$

$$\theta_j = \theta_j + (l)Err_j$$

Thuật toán sẽ tiếp tục lặp lại bước 2 cho đến khi thỏa điều kiện dừng và cho ra các tập trọng số và độ lệch tốt nhất.

## 2.2 Mạng neural tích chập (Convolutional neural network)

### 2.2.1 Giới thiệu mạng neural tích chập

Trong vài năm trở lại đây, chúng ta thấy được sự nở rộ của các hệ thống thông minh từ các công ty công nghệ lớn trên thế giới. Các chức năng nhận dạng, phân loại hay dự đoán được áp dụng rộng rãi vào các lĩnh vực thương mại, vận tải..v..v.

Mô hình Deep learning được sử dụng phổ biến và phát triển giúp các hệ thống thông minh có độ chính xác cao ngày nay chính là Convolutional Neural Networks(CNN) - mạng neuron tích chập. Trong các nội dung tới, chúng ta sẽ tìm hiểu các khái niệm, kiến trúc, cũng như ứng dụng của CNN trong lĩnh vực phân loại ảnh.

### 2.2.2 Mô hình mạng neural tích chập

#### Input và output

Phân loại ảnh là công đoạn chuyển hóa từ một đầu là một hình ảnh và kết quả là một nhãn ứng với hình ảnh đó hoặc là các xác suất mà hệ thống dự đoán dựa trên đặc điểm của ảnh. Với con người, công việc nhận diện này được hình thành từ khi mới sinh ra, chúng ta có thể đưa ra kết quả của một hình ảnh bất kỳ mà không chút khó khăn. Nhưng máy tính thì không đơn giản như vậy, đầu vào và kết quả phải được đưa về dạng kỹ thuật số mà máy có thể hiểu được.

Khi một máy tính nhận vào một ảnh, nó sẽ thấy một mảng các giá trị

pixel tùy thuộc vào kích thước và độ phân giải của ảnh[3]. Ví dụ, một ảnh màu có kích thước  $224 \times 224$  pixel thì máy tính sẽ thấy hình ảnh này dưới dạng một mảng có kích thước  $224 \times 224 \times 3$ , giá trị 3 do thuộc tính ảnh màu(RGB) mà có được, giá trị này sẽ là 1 nếu đây là ảnh trắng đen.

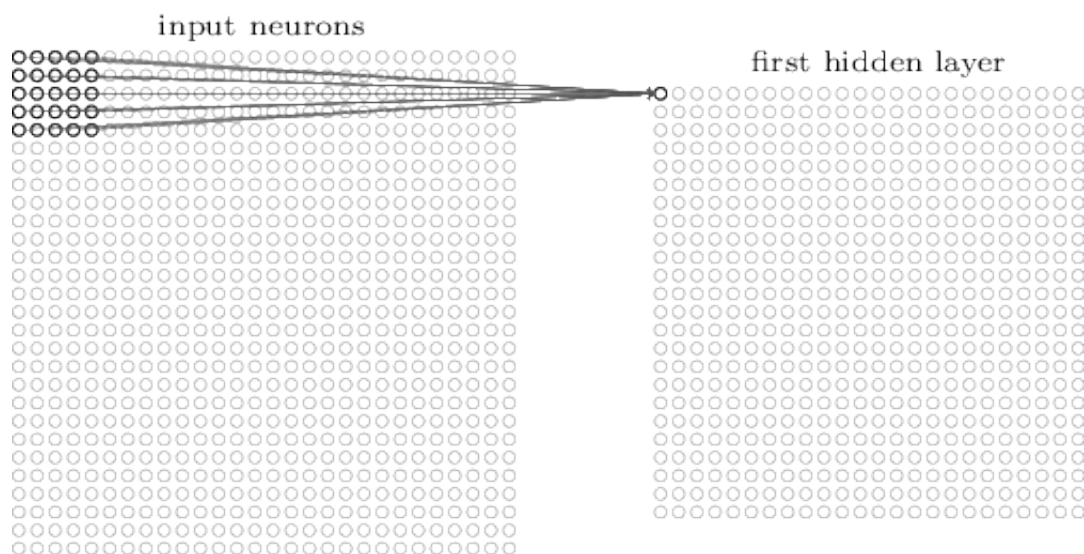
Đối với output, đây là một mảng các giá trị xác suất, mảng giá trị này cũng tùy thuộc vào số lượng nhãn(lớp) cần dự đoán. Ví dụ, (0.90 cho xe ô tô, 0.1 cho xe tải).

## Convolutional layer

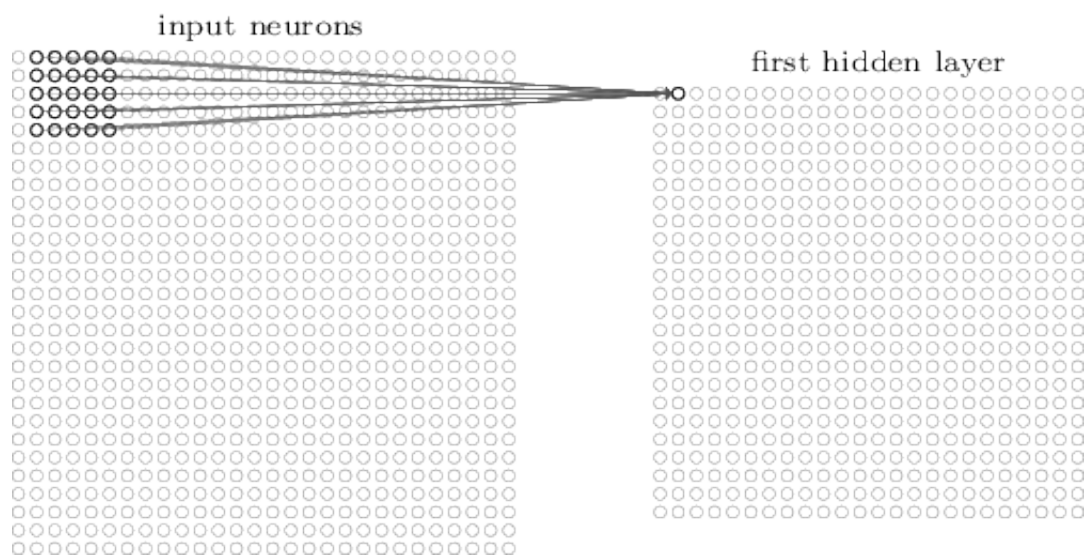
Tầng đầu tiên của một mạng CNN luôn luôn là tầng tích chập (convolutional layer)[4]. Như đã biết đầu vào (input) là một mảng các giá trị pixel. Trong trường hợp cụ thể để dễ hình dung ta chọn input là mảng các giá trị pixel có kích thước  $32 \times 32 \times 3$  với  $32 \times 32$  là chiều dài và chiều rộng của tấm hình và 3 là giá trị RGB khi là ảnh màu. Đối với input trên có nghĩa là sẽ có một ma trận có kích thước  $32 \times 32$  pixel mỗi pixel sẽ chứa 3 giá trị mà mỗi giá trị đó lần lượt biểu diễn cho giá trị của 3 màu sắc trên máy tính là đỏ(RED), lục(GREEN) và lam(BLUE).

Tạm thời bỏ qua giá trị RGB để đi vào cách hoạt động của tầng tích chập này. Cách đơn giản để giải thích cách hoạt động của tầng tích chập là tưởng tượng sẽ có một khuôn sẽ trượt từ phía trên bên trái cho đến hết tấm ảnh[5]. Với kích thước ảnh là  $32 \times 32$  như trên, chọn kích thước ô trượt ví dụ là  $5 \times 5$ . Ô trượt có kích thước  $5 \times 5$  sẽ trượt lần lượt qua cả input ảnh, ô trượt này được gọi là kernel hay filter(bộ lọc). Bộ lọc là một mảng các giá trị trọng số. Một điểm ghi chú là chiều sâu của bộ lọc sẽ bằng với chiều sâu của ảnh, với input  $32 \times 32 \times 3$  thì bộ lọc cũng sẽ có  $5 \times 5 \times 3$ . Hình 2.8 và

2.9 minh họa cách kernel trượt trên input.

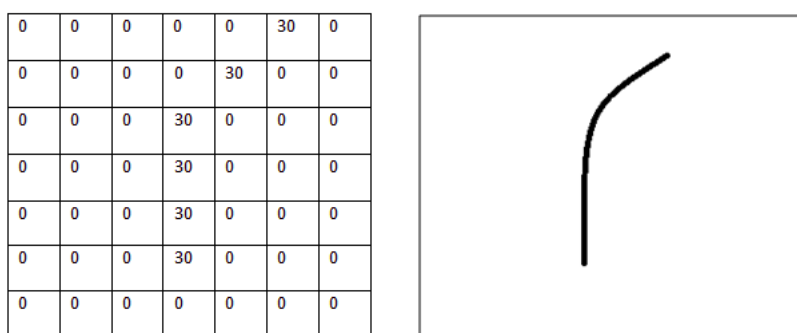


Hình 2.8: Convolutional layer



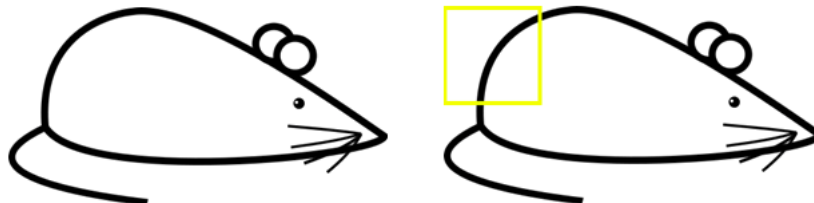
Hình 2.9: Convolutional layer

Bây giờ ta sẽ đi vào việc mà tăng tích chập thực sự làm với những phép tính. Như đã trình bày rằng mỗi bộ lọc sẽ là một mảng các giá trị pixel, công dụng của mảng giá trị này nhằm mục đích phát hiện các đặc tính của mỗi vùng input mà filter trượt qua. Các đặc tính ở đây có thể là đường thẳng, đường cong, màu đơn giản. Ví dụ ta có một filter có kích thước là  $7 \times 7 \times 3$  dùng để phát hiện một dạng đường cong như hình 2.10.

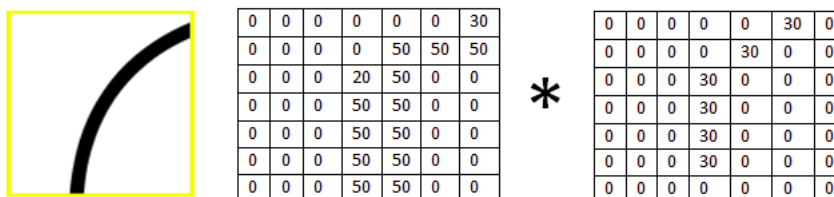


Hình 2.10: Mảng các giá trị của bộ lọc

Khi bộ lọc trên trượt đến vùng được đánh dấu vàng có dạng đường cong giống với bộ lọc như hình 2.11. Lúc đó phép toán tích chập sẽ được thực hiện như hình 2.12 với ma trận bên trái chính là giá trị pixel của vùng được đánh dấu trên ảnh mà bộ lọc trượt tới, ma trận bên phải chính là bộ lọc được sử dụng hiện tại.



Hình 2.11: ảnh đầu vào



Hình 2.12: phép toán tích chập

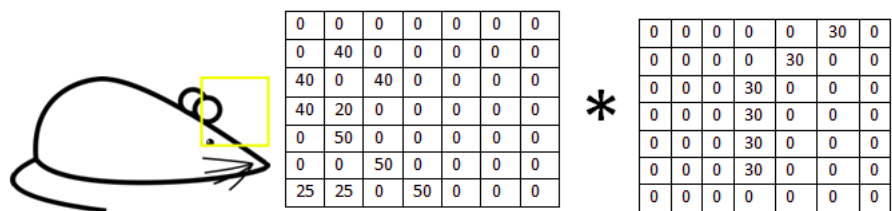
Kết quả của phép tính trên sẽ có kết quả như sau:

$$(50 * 30) + (50 * 30) + (50 * 30) + (20 * 30) + (50 * 30) = 6600$$

Đây là một con số rất lớn, thông thường nếu bộ lọc trượt tới một vùng mà vùng đó có hình dạng tương tự như bộ lọc thì kết quả khi thực hiện phép tính là một con số rất lớn. Ngược lại, kết quả sẽ ra rất nhỏ hoặc bằng 0. Ví dụ là hình ảnh 2.13 kết quả sẽ là 0 khi thực hiện phép tính

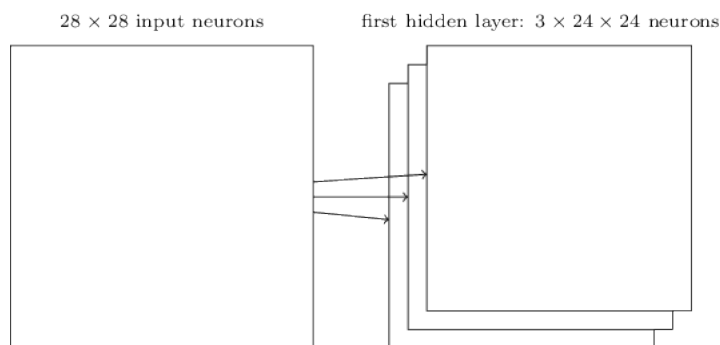
Trên đây là mô tả đơn giản về tầng tích chập, trên thực tế, chúng ta có thể có nhiều bộ lọc để phát hiện các khuôn mẫu, hình dạng khác nhau trong ảnh đầu vào. Kết quả xuất của tầng tích chập thứ nhất còn được gọi là bản





Hình 2.13: phép toán tích chập

đồ đặc tính (feature map) và có thể có nhiều feature map cho một input sau khi hoàn thành tầng tích chập. Như 2.14 biểu diễn một input kích thước như hình sau khi qua tầng tích chập và sử dụng tập các bộ lọc 5 x 5 cho ra tập 3 feature map mà mỗi cái nhận diện được một khuôn dạng khác nhau xuất hiện trong input.

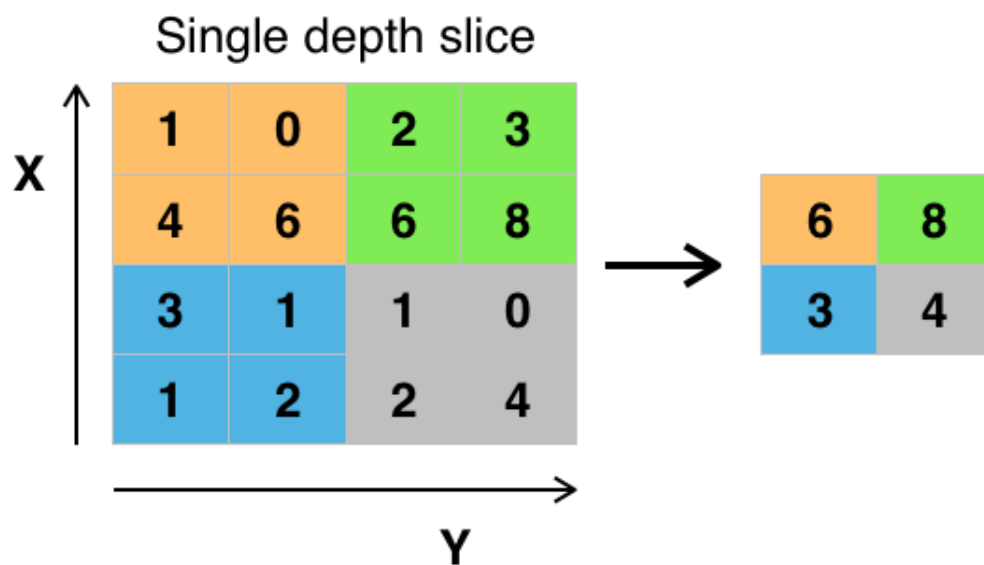


Hình 2.14: kết quả tầng tích chập

## Pooling Layer

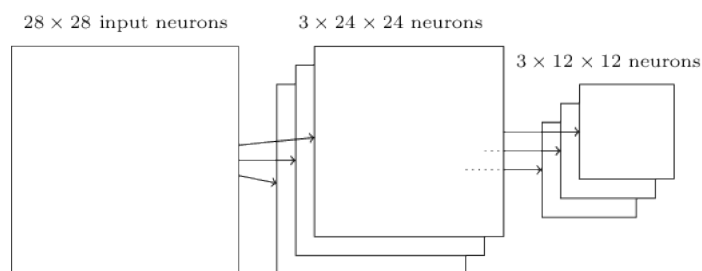
Sau khi qua một hoặc vài tầng tích chập, CNN sẽ chứa các tầng tổng hợp (Pooling Layer) ngay sau đó. Ở đây, tầng tổng hợp sẽ đơn giản hóa thông tin được lấy từ đầu ra từ tầng tích chập trước đó. Có nhiều kiểu tổng hợp

khác nhau đối với tầng này, nhưng max-pooling là phép tổng hợp phổ biến được sử dụng. Hình 2.15 biểu diễn một ví dụ phép max-pooling với một bộ lọc có kích thước là  $2 \times 2$ . Giá trị lớn nhất trong mỗi vùng được trượt qua sẽ được chọn làm kết quả xuất ra.



Hình 2.15: max-pooling  $2 \times 2$

Công dụng của phép pooling này giúp giảm đi kích thước của tập miêu tả đặc trưng từ đó cũng làm cho số lượng tham số và tính toán giảm theo. Và do chúng ta có thể có nhiều feature map từ tầng tích chập nên phép tổng hợp cũng sẽ được áp dụng độc lập cho mỗi feature map. Nếu có 3 feature map thì sẽ có 3 phép tổng hợp trong trường hợp này là max-pooling.



Hình 2.16: Ví dụ tầng tổng hợp

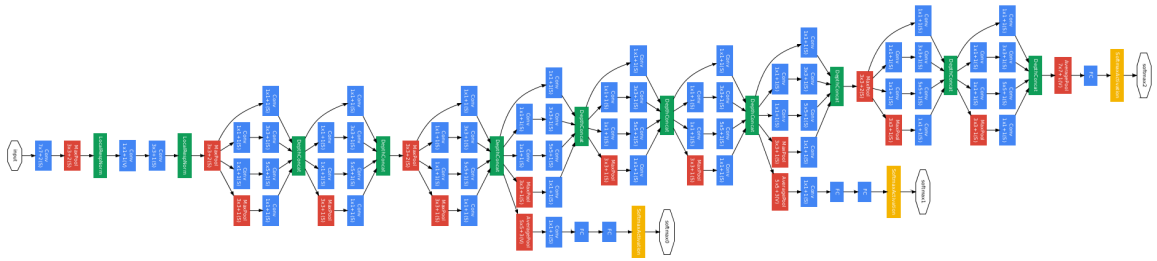
## Fully Connected Layer

Sau khi thông tin input đã qua các tầng tích chập và tổng hợp, các giá trị dữ liệu sẽ đi đến tầng fully connected để xuất ra kết quả. Kết quả tại tầng fully connected là một vector có kích thước bằng với số lớp(nhãn) mà bài toán cần dự đoán. Như bài toán phân loại ảnh giao thông ùn tắc và thông thoáng thì lúc này số nhãn cần dự đoán và kích thước vector tại tầng này sẽ bằng 2. Giá trị của các phần tử trong vector sẽ là giá trị xác suất của mỗi nhãn mà mạng dự đoán,  $[0.8, 0.2]$  sẽ biểu diễn 80% ảnh này thuộc lớp 1 và 20% ảnh thuộc lớp thứ 2. Về cơ bản, cách kết nối ở tầng này giống như cách kết nối neuron giữa các tầng với nhau ở mạng neuron ở mục trước. Khi đó, tất cả neuron ở tầng pooling sẽ kết nối với từng neuron trong tầng cuối.

### 2.2.3 Kiến trúc mạng GoogLeNet)

#### Ý tưởng

Đây là kiến trúc mạng tích chập với 22 tầng. GoogLeNet còn là quán quân của ILSVRC 2014 [9]. Mạng googLeNet có cấu trúc mạng nằm trong mạng, có 9 tầng mà mỗi tầng là một inception module. Theo tài liệu cho biết, việc áp dụng inception module giúp làm giảm đáng kể số lượng tham số tính toán giúp giải quyết vấn đề về tài nguyên. 2.17 minh họa cho cấu trúc của mạng googLeNet.

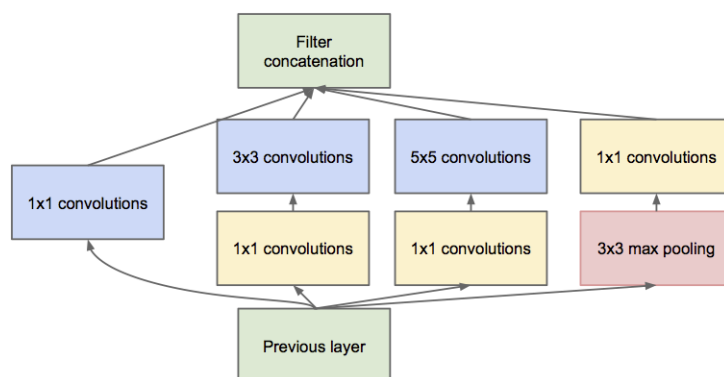


Hình 2.17: GoogLeNet

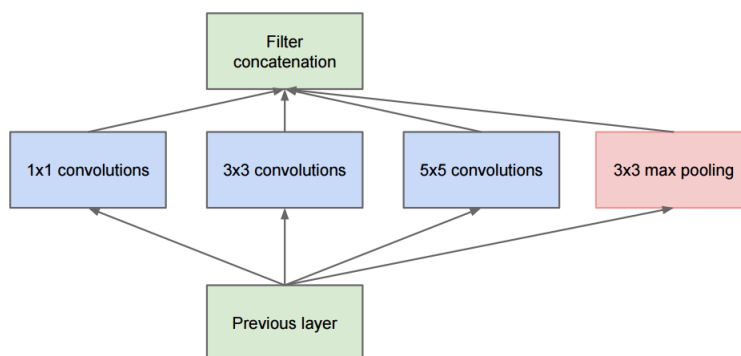
#### Inception module

Với minh họa kiến trúc của mạng googLeNet, ta sẽ thấy các layer là một khối mạng nhỏ nằm bên trong. Đây là các inception module. Đối với các mạng tích chập thông thường, khi một tập dữ liệu bắt đầu đi vào một tầng thì sẽ chỉ có hai sự lựa chọn đó chính là tầng tích chập hoặc tầng tổng hợp, nhưng với googLeNet sẽ có tập input sẽ đi vào một lớp module tại đó sẽ các

phương thức tích chập và pooling sẽ được tính toán một cách song song và độc lập với nhau [9].



Hình 2.18: inception module



Hình 2.19: naive inception module

Hình 2.18 miêu tả cấu trúc của một module trong mạng. Trong khi đó 2.19 là một ý tưởng ban đầu mà tác giả đã nghĩ tới. Ở 2.18 chúng ta thấy trước khi thực hiện các phép tích chập với các filter 3 x 3 và 5 x 5, input đều được xử lý qua phép tích chập với filter 1 x 1. Các bộ lọc 1 x 1 có tác dụng làm giảm đi chiều của các input[8], điều này giúp cho khối lượng tham số

phải tính toán ở phép toán tích chập với bộ lọc  $3 \times 3$  và  $5 \times 5$  sẽ được giảm đi một cách đáng kể.

## Chương 3

# Apache Hadoop

### 3.1 Tổng quan về hệ thống Apache Hadoop

Apache Hadoop là một dự án phát triển phần mềm nhằm cung cấp một nền tảng phân tán, có thể mở rộng linh hoạt và có độ tin cậy cao.

Ngoài ra Hadoop còn được xem là một thư viện hay framework cho phép xử lý phân tán khối lượng lớn dữ liệu trên nhiều cụm máy tính bằng các mô hình lập trình. Framework này được thiết kế với mục đích có khả năng mở rộng từ một máy chủ đơn lẻ lên đến rất nhiều trạm làm việc mà mỗi máy trạm có khả năng tính toán và lưu trữ cục bộ.

## 3.2 Hadoop distributed file system - HDFS

### 3.2.1 Thiết kế

HDFS là một hệ thống file nhằm lưu trữ một lượng rất rất lớn (Lớn ở đây theo nghĩa có thể là hàng trăm megabytes, gigabytes, hoặc terabytes) với cơ chế streaming data access trên những thiết bị phổ thông <sup>1</sup>. Đây là cơ chế phân luồng dữ liệu trong HDFS với mục đích ghi một lần chạy nhiều lần. Điển hình là dữ liệu sẽ được sinh và sao chép từ nguồn và sau đó có rất nhiều tiến trình phân tích khác nhau thực thi dữ liệu trên. Mỗi hoạt động phân tích sẽ thực thi liên quan đến một phần nào đó khác nhau trên cả một tập dữ liệu trên.

Từ các thiết bị phổ thông ở đây là những những thiết bị phân cứng máy tính, HDFS không yêu cầu một phần cứng đắt tiền hay độ tin cậy cao. Mà nó được thiết kế để chạy trên những cụm máy tính từ nhiều nhà cung ứng khác nhau. Vì thế mà xác suất để một node(đơn vị phần cứng) gặp lỗi và thất bại là rất lớn, đặc biệt là những cụm có hàng ngàn máy trạm. Với đặc tính đó, HDFS được thiết kế sao cho không có sự gián đoạn nào được phát hiện ở người dùng khi mà việc một số lượng node gặp lỗi giữa chừng.

### 3.2.2 Ý tưởng chủ đạo

---

<sup>1</sup>các máy tính hoặc máy trạm



## Chương 4

# Mô hình loại ảnh giao thông

### 4.1 Vấn đề và mục tiêu

Đối với vấn đề phát hiện kẹt xe qua hình ảnh camera, chúng ta sẽ sử dụng mạng neuron tích chập với tập ảnh huấn luyện được trích xuất từ những đoạn video do camera hành trình xe buýt ghi lại trong thời gian hoạt động. Ngoài ra, chúng ta còn sử dụng mô hình googLeNet để áp dụng vào vấn đề và phương pháp transfer learning (sẽ được trình bày trong mục tiếp theo).

### 4.2 Phương pháp transfer learning

Ngày nay, với sự phát triển của Deep learning do nguồn dữ liệu to lớn và các máy tính ngày càng cải tiến về khả năng tính toán khiến cho kết quả chính xác của các bài toán phân loại ngày càng cao. Như mô hình mạng googLeNet có rất nhiều tầng khiến cho việc huấn luyện tốn kém thời gian. Thay vì như vậy chúng ta áp dụng phương pháp transfer learning.

Transfer learning là công đoạn lấy model đã được huấn luyện từ một tập dữ liệu khác [6] và tích hợp lại với tập dữ liệu có đang có. Việc này là khả thi do mô hình đã huấn luyện trước đó sử dụng tập ảnh khổng lồ giúp cho mô hình học được loạt những đặc tính thường thấy trong cùng một ảnh. Có thể thấy tất cả các ảnh đều có những đặc tính cơ bản giống nhau, khi muốn huấn luyện lại cho tập ảnh của chính mình chúng ta chỉ cần thay tầng cuối cùng của mạng bằng tập dữ liệu của mình. Mô hình sẽ tự điều chỉnh lại các trọng số cũng như độ lệch từ các giá trị từ mô hình đã huấn luyện với tập dữ liệu mới mà không cần làm lại từ đầu.

## 4.3 Hiện thực

### 4.3.1 Tổng quan về tập dữ liệu

Để xây dựng mô hình phân loại hình ảnh, chúng ta cần phải có một tập huấn luyện đủ tốt. Ở đây, các hình ảnh được trích xuất từ camera hành trình từ các tuyến xe buýt.

Cấu trúc tổ chức tập dữ liệu gồm 2 thư mục chính:

- Thư mục **ket**: chứa các hình ảnh được cho là giao thông trong tình trạng ùn tắc. Bao gồm 1077 hình ảnh định dạng JPG.
- Thư mục **thong**: chứa các hình ảnh được cho là giao thông trong tình trạng thông thoáng. Bao gồm 2000 hình ảnh định dạng JPG.

### 4.3.2 Môi trường

Bộ phân loại ảnh giao hông được huấn luyện trên nền tảng hệ điều hành Linux, ngôn ngữ Python phiên bản 3.6 kết hợp với thư viện Tensorflow mã nguồn mở chuyên được sử dụng cho những mô hình học sâu.

Tensorflow[11] là một thư viện học sâu mã nguồn mở được Google phát triển. Thư viện này đã thu hút được sự chú ý lớn từ cộng đồng Deep-learning. Tensorflow cho phép chạy các thuật toán machine learning trên nhiều GPU, có nhiều module được dựng sẵn giúp cho việc xây dựng và thực thi mô hình đơn giản hơn.

#### Các khái niệm:

- **Tensor:** đây là cấu trúc dữ liệu được sử dụng hoàn toàn trong Tensorflow. Hay nói cách khác, tất cả dữ liệu đều biểu diễn dưới dạng tensor. Đơn giản, tensor là một mảng gồm n chiều hay list kèm theo một số thuộc tính khác.
- **Rank:** còn được gọi là số chiều của dữ liệu

Rank	Đơn vị số	Ví dụ
0	Scalar	$s = 123$
1	Vector	$s = [0.8, 0.1, 0.1]$
2	Matrix	$s = [[1,2,3], [4,5,6], [7,8,9]]$
3	3-Tensor	$s = [ [ [1], [2], [3] ], [ [4], [5], [6] ], [ [7], [8], [9] ] ]$
n	n-Tensor	n chiều dữ liệu...

Bảng 4.1: Các chiều dữ liệu.

- **Shape:** biểu diễn chiều của tensor. Ví dụ,  $t = [[1, 2, 3], [4, 5, 6], [7, 8, 9]]$  có shape là  $[3, 3]$ ,  $t = [[[1], [2], [3]], [[4], [5], [6]], [[7], [8], [9]]]$  có shape là  $[1, 3, 3], \dots$
- **Type:** là các kiểu dữ liệu được sử dụng trong Tensorflow. Một vài kiểu dữ liệu cơ bản như.

Data type	Python code	Mô tả
<i>DT-FLOAT</i>	tf.float32	32 bits floating point.
<i>DT-DOUBLE</i>	tf.float64	64 bits floating point.
<i>DT-INT16</i>	tf.int16	16 bits signed integer.
<i>DT-INT32</i>	tf.int32	32 bits signed integer.
<i>DT-INT64</i>	tf.int64	64 bits signed integer.
...	...	...

Bảng 4.2: Một vài kiểu dữ liệu.

## 4.4 Các bước hiện thực

### 4.4.1 Xử lý dữ liệu

Các hình ảnh được trình bày, dữ liệu được lưu vào các thư mục có chứa các tên mô tả cho đặc tính của những hình ảnh đó. Với bộ dữ liệu trên, chương trình tạo một kiểu dữ liệu dictionary với khóa chính là giá trị biểu diễn cho tên thư mục và cũng là tên class cần phân loại, value chính là đường dẫn các file ảnh tương ứng.

Để sử dụng được trong mô hình mạng, các hình ảnh sẽ được mã hóa sang một định dạng mới nhờ các phương thức hỗ trợ có sẵn trong thư viện Tensorflow. Sau khi được mã hóa, kết quả chính là các tensor có thông số shape như sau  $[299, 299, 3]$ , với hai vị trí đầu tiên chính là kích thước của hình ảnh cũng như của tensor, giá trị 3 biểu diễn cho độ sâu (ảnh màu).

Sau cùng, bộ dữ liệu đã được mã hóa được chia thành 3 tập con sử dụng với 3 mục đích khác nhau: tập huấn luyện (Training set), tập validation để tránh vấn đề overfit trong quá trình huấn luyện và tập kiểm thử dùng để kiểm tra độ chính xác của mô hình sau khi huấn luyện hoàn tất. Riêng tập huấn luyện sẽ được dùng để tạo ra các bottlenecks

### 4.4.2 Tạo các bottlenecks

Bottlenecks[10] là một từ được dùng để chỉ tầng (layer) nằm ngay trước fully-connected layer. Với kiến trúc mạng googLeNet, tầng này đã được huấn luyện tập dữ liệu trước đó nên có được kết quả đủ tốt để phân biệt được đặc tính của mỗi lớp(class) yêu cầu. Có nghĩa ở bước này chúng ta sẽ tạo ra một bản tóm tắt các giá trị trọng số đủ tốt cho mỗi ảnh input. Tầng cuối cùng

của kiến trúc mạng sẽ sử dụng các giá trị bottlenecks này để huấn luyện và điều chỉnh để phân loại các lớp mới. Điều này nhờ vào việc mạng đã được huấn luyện bởi tập dữ liệu gồm 1000 lớp khác nhau của ImageNet giúp cho việc phát hiện các mẫu đặc tính trở nên dễ dàng hơn.

### 4.4.3 Huấn luyện

Sau khi hoàn tất tạo các giá trị bottleneck, việc thực hiện thực hiện cấu hình mạng và huấn luyện bắt đầu. Tổng số bước huấn luyện sẽ được cài đặt mặc định là 4000 bước, tuy nhiên có thể thay đổi lại tùy theo tình huống. Mỗi bước huấn luyện sẽ chọn ra 100 dữ liệu ngẫu nhiên <sup>1</sup> để đưa vào tầng cuối cùng <sup>2</sup> để dự đoán lớp, lớp dự đoán sẽ được so sánh với các lớp thực tế để mạng điều chỉnh và cập nhật các giá trị trọng số thông qua cơ chế lan truyền ngược như đã trình bày ở chương trước. Do phép toán được thực hiện trên tập huấn luyện nên sẽ gây ra vấn đề overfit, vì thế mà tập validation sẽ được sử dụng để đo lại giá trị sai lệch và độ chính xác. Nếu độ chính xác tại tập huấn luyện cao nhưng tại tập validation không thay đổi hoặc thấp thì chứng tỏ mô hình mạng gặp phải vấn đề overfit và việc huấn luyện tiếp tục không còn có ích.

---

<sup>1</sup>Tập dữ liệu lúc này là những bottlenecks

<sup>2</sup>Tầng fully-connected với softmax là activation function

## Chương 5

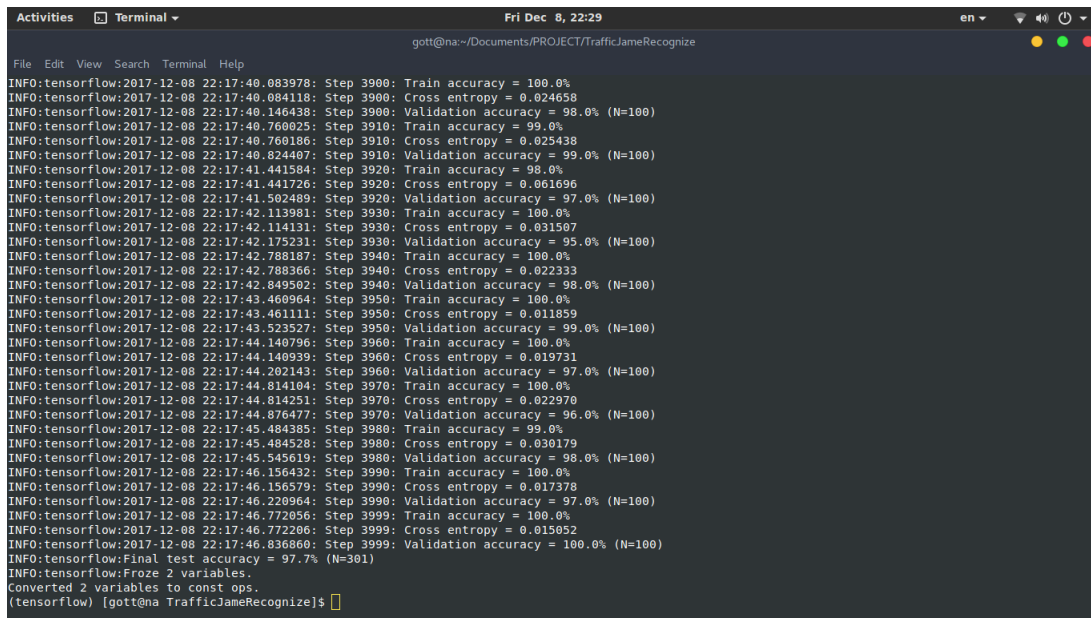
# Kết quả huấn luyện và đánh giá

### 5.1 Kết quả huấn luyện

Với các thông số cầ mạng đã cấu hình ở chương trước, 5.1 cho thấy kết quả huấn sau khi huấn luyện mạng có kết quả tương đối tốt với giá trị  $cross\_entropy = 0.015052$  và độ chính xác ở tập kiểm thử là 97.7%. Không những thế việc áp dụng kỹ thuật transfer learning giúp tối tiết kiệm thời gian huấn luyện gấp nhiều lần. Cụ thể, việc xây dựng mô hình thủ công và huấn luyện lại từ đầu đối với máy tính không hỗ trợ card GPU sẽ mất thời gian là 8 tiếng nhưng đối với việc áp dụng mô hình inception model với kỹ thuật transfer learning thì chỉ mất từ 45 đến 60 phút để hoàn thành.

### 5.2 Sử dụng mô hình phân loại ảnh mới

Sử dụng mô hình để phân loại một số hình ảnh khác chưa được phân loại. 5.2 và 5.3 có thể nhận biết bằng mắt thường đây là hình ảnh của những



```
Activities Terminal Fri Dec 8, 22:29
gott@na:~/Documents/PROJECT/TrafficJameRecognize

File Edit View Search Terminal Help

INFO:tensorflow:2017-12-08 22:17:40.083978: Step 3900: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:40.084118: Step 3900: Cross entropy = 0.024658
INFO:tensorflow:2017-12-08 22:17:40.146438: Step 3900: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:40.760025: Step 3910: Train accuracy = 99.0%
INFO:tensorflow:2017-12-08 22:17:40.760186: Step 3910: Cross entropy = 0.025438
INFO:tensorflow:2017-12-08 22:17:40.824407: Step 3910: Validation accuracy = 99.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:41.441504: Step 3920: Train accuracy = 99.0%
INFO:tensorflow:2017-12-08 22:17:41.441726: Step 3920: Cross entropy = 0.061696
INFO:tensorflow:2017-12-08 22:17:41.502489: Step 3920: Validation accuracy = 97.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:42.113981: Step 3930: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:42.114131: Step 3930: Cross entropy = 0.031507
INFO:tensorflow:2017-12-08 22:17:42.175231: Step 3930: Validation accuracy = 95.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:42.788187: Step 3940: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:42.788366: Step 3940: Cross entropy = 0.022333
INFO:tensorflow:2017-12-08 22:17:42.849502: Step 3940: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:43.460964: Step 3950: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:43.461111: Step 3950: Cross entropy = 0.011859
INFO:tensorflow:2017-12-08 22:17:43.523527: Step 3950: Validation accuracy = 99.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:44.140796: Step 3960: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:44.140939: Step 3960: Cross entropy = 0.019731
INFO:tensorflow:2017-12-08 22:17:44.202143: Step 3960: Validation accuracy = 97.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:44.814104: Step 3970: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:44.814251: Step 3970: Cross entropy = 0.022970
INFO:tensorflow:2017-12-08 22:17:44.876477: Step 3970: Validation accuracy = 96.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:45.484385: Step 3980: Train accuracy = 99.0%
INFO:tensorflow:2017-12-08 22:17:45.484528: Step 3980: Cross entropy = 0.030179
INFO:tensorflow:2017-12-08 22:17:45.545619: Step 3980: Validation accuracy = 98.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:46.156432: Step 3990: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:46.156579: Step 3990: Cross entropy = 0.017378
INFO:tensorflow:2017-12-08 22:17:46.220964: Step 3990: Validation accuracy = 97.0% (N=100)
INFO:tensorflow:2017-12-08 22:17:46.772056: Step 3999: Train accuracy = 100.0%
INFO:tensorflow:2017-12-08 22:17:46.772206: Step 3999: Cross entropy = 0.015052
INFO:tensorflow:2017-12-08 22:17:46.836860: Step 3999: Validation accuracy = 100.0% (N=100)
INFO:tensorflow:Final test accuracy = 97.7% (N=301)
INFO:tensorflow:Froze 2 variables.
Converted 2 variables to const ops.
(tensorflow) [gott@na TrafficJameRecognize]$
```

Hình 5.1: Kết quả huấn luyện

con đường đang trong tình trạng ùn tắc. Kết quả dự đoán cho lớp cả hai hình ảnh này là chính xác theo lớp kẹt xe.





ket 0.997227  
thong 0.00277315

Hình 5.2: Ảnh kiểm thử 1 - kết quả

Đối với hai hình ảnh tiếp theo được phân biệt vào loại đường thông thoáng một cách dễ dàng. Mô hình dự đoán xác suất hai hình ảnh dưới đây thuộc lớp thông thoáng lần lượt là.



Hình 5.3: Ảnh kiểm thử 2 - kết quả

Như vậy, kết quả kiểm thử đối với một số hình ảnh chưa được phân loại của mô hình đã huấn luyện cho kết quả khá chính xác. Tuy nhiên, đối với vấn đề phân loại giao thông thì không chỉ có hai trường hợp ùn tắc hay thông thoáng mà còn tồn tại nhiều trường hợp hơn. Như tình huống đông xe nhưng di chuyển chậm, trên thực tế đây là tình huống không phải ùn tắc nhưng ảnh chụp gần giống với ảnh ùn tắc. Cần chú ý vấn đề này khi lựa chọn, phân loại ảnh huấn luyện, hoặc để giải quyết tốt hơn cần phải tăng số lượng lớp(nhãn) cần phân loại lên thành 3 hay lớn hơn thay vì 2 như ban đầu.



thong 0.719465  
ket 0.280535

Hình 5.4: Ảnh kiểm thử 3 - kết quả



thong 0.725881  
ket 0.274119

Hình 5.5: Ảnh kiểm thử 4 - kết quả

# Chương 6

## Kết luận

### 6.1 Các kết quả

Đề cương này giúp nghiên cứu, tìm hiểu về mạng học sâu cũng như việc phân loại ảnh giao thông. Các kết quả đã đạt được đáp ứng được các mục tiêu ở chương 1.

- Xây dựng được bộ dữ liệu phục vụ cho việc huấn luyện mô hình phân biệt 2 loại ảnh giao thông. Số lượng ảnh thu thập được trung bình mỗi lớp khoảng 1000 cho đến 2000 ảnh.
- Cài đặt, cấu hình các thông số cũng như mô hình mạng đã được huấn luyện trước, ứng dụng vào vấn đề phân loại ảnh giao thông.
- Kết quả huấn luyện và kiểm thử đối với mạng googLeNet thu được khá tốt.

## 6.2 Hướng phát triển

Sau khi đạt được kết quả huấn luyện khá tốt, hướng phát triển của đề tài trong tương lai như sau:

- **Bước 1.** Tiếp tục xem xét việc xác định các lớp ảnh giao thông trong tương lai giúp phát hiện nhiều loại hình như ùn tắc, thông thoáng, đông xe di chuyển chậm, v.v. giúp cụ thể hóa tình trạng giao thông.
- **Bước 2.** Xây dựng ứng dụng di động nhận biết kẹt xe.
- **Bước 3.** Phát triển Web-service nhận thông tin hình ảnh từ các camera kết nối, phối hợp với mô hình đã huấn luyện để tiến hành phân loại giao thông.

Hệ thống tiềm năng trên có thể giúp dân cư sinh sống ở các khu đô thị cũng như ban quản lý nắm bắt tình hình giao thông để phân luồng di chuyển và khắc phục một cách nhanh nhất.

# Tài liệu tham khảo

- [1] Rémi Cadène, Nicolas Thome, and Matthieu Cord. Master's Thesis : Deep Learning for Visual Recognition. 2016.
- [2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. Chapter 6:Deep Feedforward Networks.
- [3] Andrej Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. CNN overview, 2016.
- [4] Andrej Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. convolutional layers, 2016.
- [5] Andrej Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. CNN architecture, 2016.
- [6] Andrej Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. transfer learning, 2016.
- [7] Andrej Karpathy. CS231n: Convolutional Neural Networks for Visual Recognition. module 1 - Neural Networks and module 2 - Convolutional Neural Network, 2016.

- [8] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. 2013.
- [9] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed and Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. 2014.
- [10] tensorflow.org. Image retraining.
- [11] tensorflow.org. Tensorflow programmer guide.
- [12] Matthew D. Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. 2013.