

ĐỒ ÁN TỐT NGHIỆP

Đề tài: Dự đoán mật độ giao thông bằng phương pháp Học sâu

**Giảng viên hướng dẫn:** Nguyễn Quang Trung

**Người thực hiện**: Phạm Đức Thắng

**MỤC LỤC**

# 

# **LỜI CẢM ƠN**

Trong suốt quá trình học tập, nghiên cứu và thực hiện đồ án, tôi đã luôn nhận được sự dạy dỗ, chỉ bảo tận tình của các thầy, cô giáo cũng như sự quan tâm chăm sóc từ gia đình và bạn bè. Đó là nguồn động lực giúp tôi vượt qua mọi khó khăn để hoàn thành đồ án này.

Đặc biệt, cho phép tôi được bày tỏ lòng biết ơn sâu sắc tới TS. Nguyễn Quang Trung đã dìu dắt, hướng dẫn, giúp tôi hòa nhập với môi trường nghiên cứu khoa học, tạo điều kiện thuận lợi nhất cho tôi trong suốt quá trình làm thực nghiệm. Qua đó tôi đã biết cách tiếp cận một nghiên cứu khoa học và có hành trang vững chắc cho con đường sau này.

Cuối cùng, tôi xin gửi lời cảm ơn tới các anh chị cán bộ trong, gia đình, người thân và các bạn bè luôn ở bên giúp đỡ, động viên tôi vượt qua những khó khăn trong học tập suốt những năm vừa qua.

Tôi xin chân thành cảm ơn!

Hà Nội, ngày tháng năm 2023

Sinh viên

Phạm Đức Thắng

# **ĐẶT VẤN ĐỀ**

Trí tuệ nhân tạo (Artificial Intelligence – viết tắt là AI) là một ngành thuộc lĩnh vực Khoa học máy tính. Thuật ngữ AI thường được sử dụng để chỉ các máy móc, máy tính có khả năng bắt chước được chức năng “nhận thức” của con người. Theo đó, Thị giác máy tính là một nhánh trong AI, với mục đích là “dạy” cho máy tính có thể hiểu và cảm nhận được môi trường xung quanh thông qua hình ảnh số thu được từ máy ảnh, máy quay,…

Các tác vụ cơ bản thường thấy trong Thị giác máy tính là Phân loại hình ảnh (image classification), Định vị vật thể (object localization), Nhận diện đối tượng (object detection), Phân mảnh hình ảnh (image segmentation), Mô tả hình ảnh (image captioning).

Trong thời đại chuyển đổi số, công nghệ AI được ứng dụng vào nhiều lĩnh vực trong đời sống hiện đại. Đối với ngành giao thông vận tải, bài toán phân luồng giao thông để tối ưu thời gian xe cộ di chuyển trên đường, giảm ách tắc ùn ứ vẫn là một thách thức đối với cơ quan chức năng. Nhằm hỗ trợ đội ngũ cán bộ thực hiện nhiệm vụ trong ngành giao thông, Thị giác máy tính đã được áp dụng và phát triển các thuật toán để giải quyết nhiều vấn đề, có thể kể đến như Nhận diện biển số xe, Xử lý phạt nguội, v.v… Trong khóa luận này, tác giả xin trình bày giải pháp sử dụng Thị giác máy tính để dự đoán Mật độ giao thông từ một tập hợp các ảnh đầu vào và/hoặc video.

Đo lường và dự đoán Mật độ giao thông là một bài toán nhỏ thuộc dạng Phân loại hình ảnh (image classification). Đây là bài toán được nhiều nhà nghiên cứu quan tâm bởi tính thiết thực của nó. Bằng việc ứng dụng kỹ thuật Học chuyển giao (transfer learning) để huấn luyện các mô hình học sâu như AlexNet, LeNet-5 và một số mô hình được cung cấp bởi thư viện Tensorflow, khóa luận này báo cáo kết quả của thí nghiệm dự đoán Mật độ giao thông tự động được đánh giá thông qua các chỉ số đo như f-score, độ chính xác (accuracy) đạt lần lượt trên 0.85 và 80%, cao hơn so với mô hình học máy cổ điển.

**Từ khóa:***Đo lường, dự đoán, Mật độ giao thông, tự động, Thị giác máy tính, học chuyển giao, mô hình Học sâu, chính xác.*

**Chương 1: Giới thiệu**

**1.1** **Lý do chọn đề tài**

Phân tích và đo lường mật độ giao thông là một lĩnh vực quan trọng trong ngành giao thông vận tải. Với mục đích giúp cải thiện kế hoạch vận chuyển và giảm thiểu ùn tắc giao thông, các chuyên gia sử dụng phương pháp phân tích hình dạng để xác định mật độ giao thông trong khu vực nhất định. Phân tích hình dạng giúp các chuyên gia hiểu rõ hơn về đa dạng các loại phương tiện và khối lượng lưu thông trên đường, từ đó đưa ra các giải pháp để tối ưu hóa quản lý giao thông.

Trong Thị giác máy tính, phân loại hình ảnh là một bài toán con quan trọng thuộc bài toán phân tích hình dạng. Hệ thống phân loại phải xác định được nhãn (label) phù hợp nhất cho dữ liệu đầu vào như ảnh hoặc/và video. Ứng dụng thường gặp nhất của hệ thống này trong thực tế là phân tích dữ liệu từ camera giám sát đường phố để xác định mật độ giao thông và đưa ra các giải pháp phù hợp.

Thật vậy, mật độ giao thông luôn là chủ đề rất quan trọng được nghiên cứu và quan tâm trong việc phát triển các giải pháp thông minh cho giao thông đô thị. Việc đo lường mật độ giao thông có thể giúp chúng ta đánh giá hiệu quả của hệ thống giao thông, đồng thời cung cấp thông tin quan trọng để đưa ra các quyết định cải tiến cơ sở hạ tầng.

Như đã trình bày tại phần Đặt vấn đề, bài toán đo mật độ giao thông, các nghiên cứu gần đây đã áp dụng các kỹ thuật Học máy và Học sâu. Các phương pháp Học máy truyền thống như (Support Vector Machine) SVM, (Random Forest) RF đã được sử dụng để phân loại và dự đoán mật độ giao thông trên các tuyến đường khác nhau. Các phương pháp Học sâu mang lại kết quả chính xác hơn với việc sử dụng các mô hình mạng nơ-ron sâu, cho phép phân tích các đặc trưng phức tạp từ dữ liệu ảnh giao thông.

**1.2** **Các nghiên cứu liên quan**

Thực tế đã có một số nghiên cứu làm về phân loại mật độ giao thông như bài báo “Traffic Density Classification for Multiclass Vehicles Using Customized Convolutional Neural Network for Smart City” của Deepak Mane và cộng sự Nihar Ranjan, “Classification of Traffic Vehicle Density Using Deep Learning” của Abdul Kholik cùng 2 người công sự Agus Harjoko và Wahyono Wahyono, hay “An image generation approach for traffic density classification at large-scale road network” của Jiho Cho, Hongsuk Yi, Heejin Jung và Khac-Hoai Nam Bui.

Trong bài báo đầu tiên, tác giả đề xuất một phương pháp sử dụng mạng nơ-ron tích chập tùy chỉnh (CCNN) trên hình ảnh giao thông để phân loại hình ảnh theo mật độ giao thông và do đó cung cấp hỗ trợ lái xe. Hệ thống được đề xuất trong bài báo có thể giám sát giao thông bằng cách sử dụng video được quay bởi các camera được lắp đặt và sau đó phân loại tình hình giao thông hiện tại thành các loại cao, trung bình và thấp. Mục đích là sử dụng nó như một mô hình để cung cấp thông tin mật độ giao thông từ nhiều địa điểm cho các hệ thống chuyên gia và đưa ra các quyết định quan trọng khác liên quan đến kiểm soát giao thông. Đơn vị xử lý đồ họa NVIDIA (GPU) được sử dụng để song song hóa quá trình đào tạo và triển khai các mạng nơ-ron sâu phức tạp để đạt được độ chính xác tốt hơn. Đánh giá hiệu suất của hệ thống được đề xuất được thực hiện trên một tập dữ liệu thời gian thực chứa các cảnh quay giao thông từ thành phố Pune (Ấn Độ) và các cảnh quay từ các camera giám sát trên cao tốc ở Seattle, WA, thu được từ B.

Bài báo thứ hai, tác giả cho biết hiện nay, việc giám sát được tiến hành bởi cảnh sát, Sở Giao thông và các tổ chức đang sử dụng giám sát dựa trên video như CCTV vẫn được giám sát bởi con người. Deep Learning là một phương pháp của các máy học dựa trên mạng nơ-ron tổng hợp đang được phát triển và nghiên cứu tích cực gần đây vì đã thành công trong việc giải quyết các vấn đề soft-computing khác nhau. Nghiên cứu này sử dụng kiến trúc mạng nơ-ron tích chập. Nghiên cứu này cố gắng thay đổi các thông số hỗ trợ trên mạng nơ-ron tích chập để hiệu chỉnh độ chính xác tối đa. Sau khi thay đổi các thông số, mô hình phân loại được kiểm tra bằng cách sử dụng phương pháp kiểm tra chéo K-fold, ma trận nhầm lẫn và kiểm tra mô hình với dữ liệu thử nghiệm. Trên thử nghiệm kiểm tra chéo K-fold với sản lượng trung bình là 92,83% với giá trị của K (số lần chia) = 5, kiểm tra mô hình được thực hiện bằng cách nhập dữ liệu thử nghiệm lên đến 100 dữ liệu, mô hình có thể dự đoán hoặc phân loại chính xác, nghĩa là 81 dữ liệu.

Bài báo thứ ba, tác giả để xuất một phương pháp tạo hình ảnh mới để phân loại mật độ giao thông và tốc độ có thể được đo lường tại các khu vực cụ thể bằng cách sử dụng các hệ thống giám sát (ví dụ: cảm biến vòng lặp). Tuy nhiên, việc đo mật độ giao thông khó khăn và phụ thuộc vào mối tương quan không gian từ quan điểm của mạng lưới. Do đó, tác giả đưa ra phương pháp tạo hình ảnh hiệu quả, dựa trên thời gian đến và rời khỏi thông tin của các phương tiện để giải quyết vấn đề này. Đối với thực nghiệm, việc phân loại mật độ giao thông bằng một mạng nơ-ron tích chập được thực hiện trên dữ liệu thiết bị ven đường giao thông của 11 giao lộ liên tục để đánh giá tính hiệu quả của phương pháp được đề xuất.

## **1.3 Cấu trúc của đồ án**

Khóa luận được tổ chức như sau:

Chương 1: giới thiệu vấn đề và các nghiên cứu liên quan.

Chương 2: tổng quan bài toán Đo lường mật độ giao thông và phương pháp nghiên cứu chi tiết.

Chương 3: thu thập, xử lý dữ liệu và thực nghiệm với phương pháp sử dụng mạng Học sâu đã đề xuất.

Chương 4: kết quả và đánh giá các mô hình.

Chương 5: tổng kết.

**Chương 2: Kiến thức nền tảng**

**2.1. Tổng quan**

2.1.1. Phát biểu bài toán

Bài toán Đo lường mật độ giao thông có thể được phát biểu như sau: Cho một ảnh kĩ thuật số chụp một đoạn đường giao thông dưới dạng ma trận I wxh với w và h là giá trị chiều rộng và chiều cao của bức ảnh đó. Thành phần của ma trận I có thể là một giá trị nguyên thuộc [0, 255] cho ảnh đơn sắc, hoặc bộ ba giá trị nguyên thuộc [0, 255] tương ứng các màu đỏ (R), xanh lá (G), xanh dương (B) đối với ảnh màu. Hãy xác định xem tình trạng mật độ giao thông trong bức ảnh đầu vào tương ứng với mức nào trong tổng cộng 5 mức: (1) rất vắng, (2) vắng, (3) trung bình, (4) đông, (5) tắc.

Như đã trình bày tại Chương 1, mục tiêu của nghiên cứu này là phát triển một phương pháp sử dụng các mạng Học sâu để tự động hóa bài toán trên. Nhiều nghiên cứu trên thế giới đã chứng minh rằng mạng Học sâu thực sự đem lại kết quả rất tốt cho các bài toán phân loại ảnh. Đầu ra của các mạng cho bài toán phân lớp là một nhãn phù hợp nhất với nội dung ảnh đầu vào (với điều kiện là máy tính phải được huấn luyện với một tập hữu hạn các nhãn cho trước).

2.1.2. Phương pháp đánh giá mô hình

Confusion matrix

Độ chính xác (Accuracy)

**2.2. Cơ sở lý thuyết**

**2.2.1. Sơ lược về Học máy**

‘Học máy’ (Machine learning) - Một trong những thuật ngữ rất rất phổ biến trong thời đại công nghệ số hiện nay. Cách hiểu đơn giản nhất là ta dạy máy cách xử lý dữ liệu nhiều hơn một cách hiệu quả. Học máy liên quan đến những công việc liên quan đến máy tính, ta sẽ cung cấp cho máy tính khả năng học hỏi và xây dựng kỹ thuật cho phép hệ thống ‘học’ tự động từ những dữ liệu đầu vào để giải quyết vấn đề cụ thể dựa trên kinh nghiệm đưa vào.

**2.2.2. Mạng Nơ-ron**

Mạng Nơ-ron (Neural Network) là một chuỗi những thuật toán được đưa ra để tìm kiếm các mối quan hệ cơ bản trong tập hợp các dữ liệu. Thông qua việc bắt bước cách thức hoạt động từ não bộ con người. Nói cách khác, mạng nơ ron nhân tạo được xem là hệ thống của các tế bào thần kinh nhân tạo. Đây thường có thể là hữu cơ hoặc nhân tạo về bản chất.

Neural Network có khả năng thích ứng được với mọi thay đổi từ đầu vào. Do vậy, nó có thể đưa ra được mọi kết quả một cách tốt nhất có thể mà bạn không cần phải thiết kế lại những tiêu chí đầu ra. Khái niệm này có nguồn gốc từ trí tuệ nhân tạo, đang nhanh chóng trở nên phổ biến hơn trong sự phát triển của những hệ thống giao dịch điện tử.

Mạng Neural Network được cấu thành từ những tầng perceptron, gồm 3 tầng chính:

* Tầng vào (input layer): Như trong hình ở dưới, tầng này nằm bên phía trái, thể hiện cho các dữ liệu đầu vào.
* Tầng ra (output layer): Ngược lại với tầng vào, tầng ra thể hiện cho đầu ra của mạng nơ-ron và nằm bên phía bên phải của hình.
* Tầng ẩn (hidden layer): Đây là tầng nằm ở giữa, thể hiện cho quá trình xử lý thông tin và suy luận của mạng. Nó sẽ nhận các thông tin đầu vào ở đầu vào và trả kết quả ở đầu ra thông qua chức năng kích hoạt.

**2.2.3. Một số ứng dụng của mạng Nơ-ron**

* Nhận dạng chữ viết tay: tạo chữ ký online, các ký tự viết tay sẽ được chuyển đổi thành các ký tự kỹ thuật số bởi mạng nơ-ron nhân tạo.
* Nhận diện khuôn mặt, giọng nói: Dùng trong mở khóa điện thoại, hỗ trợ điều tra, xác định tội phạm, hỗ trợ pháp y.
* Tối ưu quãng đường di chuyển: Ứng dụng này nổi bật trong các phần mềm bản đồ và điển hình nhất là Google Map. Khi chúng ta tìm kiếm một địa điểm cụ thể nào đó và đường đi đến đó, Google sẽ đề xuất 2-3 con đường cho mình chọn. Trong đó, con đường ngắn nhất và tối ưu nhất sẽ được làm nổi bật.

**2.2.4. Ứng dụng trong phân loại mật độ giao thông**

Một trong những ứng dụng phổ biến của mạng neural trong phân loại mật độ giao thông là dự đoán mật độ giao thông tại các điểm trên đường và dự báo tình trạng giao thông trong tương lai. Kết quả từ mạng neural có thể được sử dụng để cung cấp thông tin thời gian thực về mật độ giao thông cho người dùng, giúp họ lựa chọn đường đi tối ưu và tránh những nơi có mật độ giao thông cao.

Quá trình ứng dụng mạng neural vào phân loại mật độ giao thông bắt đầu bằng việc thu thập dữ liệu từ các cảm biến giao thông hoặc hệ thống giám sát. Các thông tin như số lượng phương tiện, tốc độ di chuyển, đoạn đường, và thời gian trong ngày được thu thập và sử dụng làm dữ liệu đầu vào cho mạng neural.

Sau khi thu thập dữ liệu, mạng neural được huấn luyện bằng việc tối ưu hóa các trọng số của nó. Quá trình huấn luyện này dựa trên các mô hình học máy để phân tích dữ liệu đầu vào và tìm ra các mẫu, quy luật, và mối quan hệ giữa dữ liệu và mật độ giao thông. Khi đã hoàn thành quá trình huấn luyện, mạng neural có khả năng dự đoán mật độ giao thông dựa trên dữ liệu mới và không biết trước.

**2.3. Các mạng Mô hình Học sâu trong Phân loại ảnh**

**2.3.1. Một số khái niệm và thuật ngữ thường dùng**

#### 2.3.1.1 Trích xuất đặc trưng

##### a) Khái niệm:

Trong học máy, xử lý và phân loại ảnh, trích xuất đặc trưng là một khái niệm chọn lọc ra các giá trị có ý nghĩa, không rườm rà - thừa thãi từ một tập dữ liệu. Các thông tin sau khi được lựa chọn sẽ tạo điều kiện thuận lợi cho các bước học tập (của máy) và khái quát hóa. Một số trường hợp, thông qua trích xuất đặc trưng, máy tính có thể diễn giải về đối tượng được quan tâm tốt hơn con người.

##### b) Sự quan trọng của Trích xuất đặc trưng:

Trích xuất đặc trưng từ dữ liệu đầu vào giúp tăng độ chính xác của mô hình đào tạo. Giai đoạn này làm giảm số chiều của dữ liệu thông qua việc loại bỏ dữ liệu dư thừa. Điều này thực sự hữu hiệu cho trực quan hóa dữ liệu, chẳng hạn đối với một tập dữ liệu phức tạp có thể được hiển thị một cách dễ dàng khi số chiều của nó được giảm xuống còn hai hoặc ba chiều. Bởi vậy, tác vụ này tăng tốc độ huấn luyện và tốc độ suy luận của mô hình. Thông qua việc kết hợp và biến đổi tập các đặc trưng ban đầu, trích xuất đặc trưng tạo ra những đặc trưng mới có ích hơn. Các dạng đặc trưng được giữ lại thường thấy trong dữ liệu ảnh là màu sắc, hình dạng, kết cấu hoặc giá trị pixel.

##### c) Ứng dụng:

Một số ứng dụng của trích xuất đặc trưng có thể kể đến như phân tích ngữ nghĩa tiềm ẩn, nén - giảm số chiều dữ liệu, phân tích và chiếu dữ liệu, nhận diện mẫu. Như đã đề cập ở phần trên, trích xuất tính năng cũng được dùng để tăng cường tốc độ và hiệu quả của việc học có giám sát.

2.3.1.2 Phân loại ảnh:

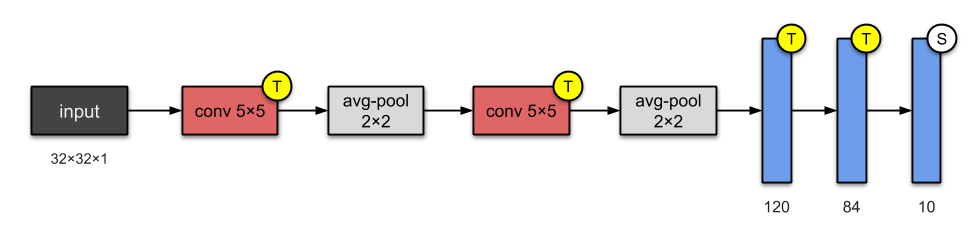
2.3.2. Cấu trúc các mạng backbone trong Học sâu

**2.3.2.1. LeNet-5**

LeNet-5 là một mô hình kiến trúc mạng nơ-ron sử dụng cho việc nhận dạng ký tự trong các ứng dụng OCR (Optical Character Recognition). Được giới thiệu bởi Yann LeCun và các đồng nghiệp vào năm 1998, LeNet-5 là một trong những mô hình đầu tiên áp dụng thành công mạng nơ-ron tích chập cho việc nhận dạng hình ảnh.

LeNet-5 gồm 5 lớp:

* Lớp tích chập (Convolutional layer): LeNet-5 sử dụng hai lớp tích chập (C1 và C3) để rút trích các đặc trưng cục bộ từ dữ liệu đầu vào. Mỗi lớp tích chập sử dụng một bộ lọc (filter) để tạo ra các đặc trưng.
* Lớp tạo mẫu (Pooling layer): Sau mỗi lớp tích chập, một lớp tạo mẫu (subsampling layer) được áp dụng để giảm kích thước của dữ liệu. Thông thường, LeNet-5 sử dụng lớp tạo mẫu theo phương pháp tạo mẫu cực đại (max-pooling) với cửa sổ 2x2 và bước nhảy (stride) là 2.
* Lớp kích hoạt (Activation layer): Sau mỗi lớp tạo mẫu, một hàm kích hoạt phi tuyến (thông thường là hàm sigmoid) được áp dụng để tạo độ không tuyến tính vào mạng nơ-ron.
* Lớp kết nối đầy đủ (Fully Connected layer): Sau hai lớp tích chập và hai lớp tạo mẫu, các đặc trưng được truyền qua hai lớp kết nối đầy đủ (F6 và F7) để ghép nối thông tin từ các lớp trước đó và chuẩn bị cho quá trình phân loại.
* Lớp đầu ra (Output layer): Lớp cuối cùng là lớp đầu ra, sử dụng hàm softmax để tính toán xác suất dự đoán cho từng lớp đối tượng có thể có trong dữ liệu.



LeNet-5 ban đầu được thiết kế để nhận dạng và phân loại chữ số viết tay. Tuy nhiên, cấu trúc và ý tưởng của LeNet-5 đã được áp dụng rộng rãi trong nhiều lĩnh vực khác nhau, bao gồm nhận dạng ký tự, nhận dạng biển số xe, nhận dạng khuôn mặt, xử lý ảnh y tế và nhiều ứng dụng khác trong lĩnh vực thị giác máy tính.

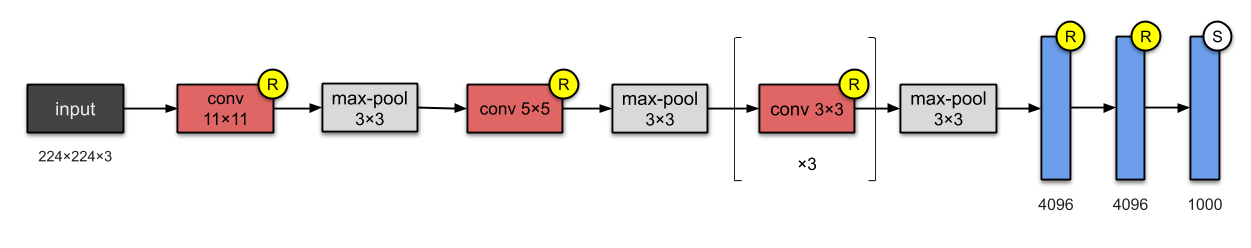
Tóm lại, LeNet-5 là một mô hình CNN quan trọng và tiền đề cho sự phát triển của các mô hình CNN hiện đại. Với cấu trúc đơn giản và hiệu suất tốt trong việc trích xuất đặc trưng từ dữ liệu hình ảnh, LeNet-5 đã tạo ra một cơ sở cho việc nghiên cứu và áp dụng thị giác máy tính trong nhiều ứng dụng thực tế.

**2.3.2.2 AlexNet**

Mô hình AlexNet là một mô hình mạng nơ-ron tích chập (CNN) tiên tiến, được giới thiệu bởi Alex Krizhevsky, Ilya Sutskever và Geoffrey Hinton vào năm 2012. Mô hình này đã đạt được thành công lớn trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2012.

Mạng AlexNet bao gồm 8 lớp:

* Lớp thứ nhất: Convolutional Layer có kích thước 11x11x3 với stride size = 4 và pad = 0. Kết quả sau bước này ta được tập feature map có kích thước 55x55x96. Tiếp theo là một Overlapping Max Pooling 3x3 có stride =2 => feature maps = 27x27x96. Tiếp theo là Local Response Normalization => feature maps = 27x27x96.
* Lớp thứ hai: Convolutional Layer: 256 kernels có kích thước 5x5x48 (stride size = 1, pad = 2) => 27x27x256 feature maps. Overlapping Max Pooling 3x3 có stride =2 => feature maps = 13x13x256. Tiếp theo là Local Response Normalization => feature maps = 13x13x256.
* Lớp thứ ba: Convolutional Layer: 384 kernels có kích thước 3x3x256 (stride size = 1, pad = 1) => 13x13x384 feature maps.
* Lớp thứ tư: 384 kernels có kích thước 3x3x192 (stride size = 1, pad = 1) => 13x13x384 feature maps.
* Lớp thứ năm: Convolutional Layer: 256 kernels có kích thước 3x3x192 (stride size = 1, pad = 1) => 13x13x256 feature maps. Overlapping Max Pooling 3x3 có stride =2 => feature maps = 6x6x256.
* Lớp thứ sáu: Full connected (hay còn gọi là Dense layer) với 4096 neurals.
* Lớp thứ bảy: Full connected (hay còn gọi là Dense layer) với 4096 neurals.
* Lớp thứ tám: Full connected ra output 1000 neurals.



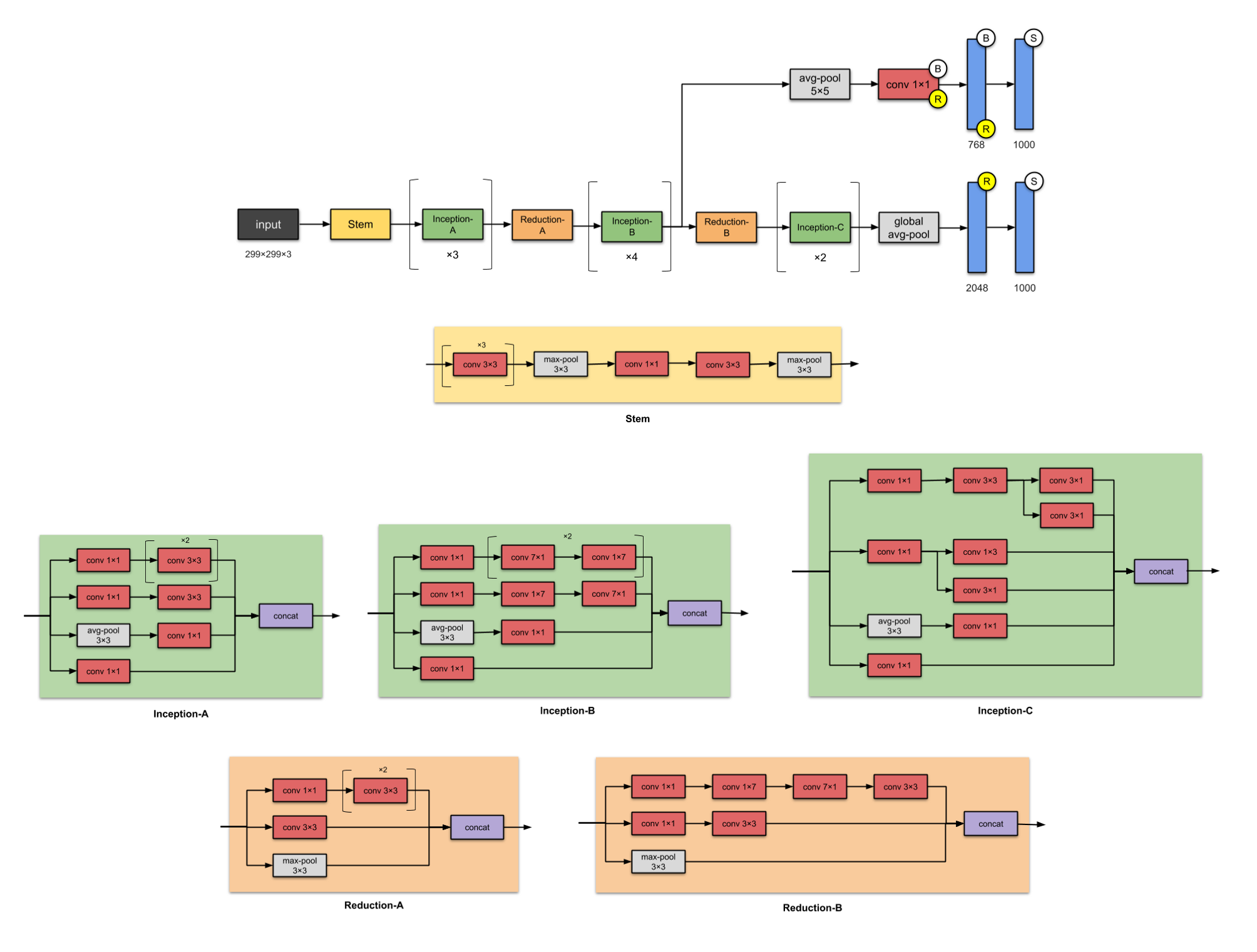
Mô hình AlexNet có một số đặc điểm nổi bật như sử dụng kỹ thuật dropout để tránh overfitting, sử dụng các kỹ thuật data augmentation để tăng cường dữ liệu huấn luyện và sử dụng hàm mất mát là cross-entropy loss để đánh giá hiệu suất của mô hình.

**2.3.2.3 InceptionV3**

Mô hình InceptionV3 là một mô hình mạng nơ-ron tích chập (CNN) được phát triển bởi Christian Szegedy và nhóm nghiên cứu tại Google vào năm 2015. Mô hình này là phiên bản tiến hóa của InceptionV1 và InceptionV2, và nó đã đạt được kết quả rất ấn tượng trong cuộc thi ImageNet Large Scale Visual Recognition Challenge (ILSVRC) năm 2015.

Cấu trúc của mô hình InceptionV3:

* Cấu trúc Inception Module: Đặc điểm nổi bật nhất của InceptionV3 là việc sử dụng Inception module, một cấu trúc được thiết kế để trích xuất các đặc trưng từ dữ liệu đầu vào ở nhiều tỷ lệ không gian khác nhau. Mỗi Inception module bao gồm việc kết hợp và xử lý đồng thời các kích thước kernel khác nhau như 1x1, 3x3 và 5x5 để tạo ra các đặc trưng có ý nghĩa từng phần riêng biệt. Kết quả của các lớp tích chập này được kết hợp thông qua phép nối (concatenation) để tạo ra đầu ra cuối cùng của mỗi Inception module.
* Lớp Average Pooling Layer: Sau khi trích xuất các đặc trưng bằng các Inception module, InceptionV3 sử dụng lớp average pooling để giảm kích thước không gian của đặc trưng và tạo ra một vector đặc trưng tổng quát.
* Lớp Fully Connected Layer: Sau lớp pooling, InceptionV3 sử dụng một hoặc nhiều lớp fully connected để kết nối các đặc trưng đã trích xuất với các lớp ẩn và lớp đầu ra. Các lớp fully connected thường được sử dụng để phân loại hoặc dự đoán.
* Lớp Softmax: Cuối cùng, InceptionV3 sử dụng lớp softmax để đưa ra các xác suất phân loại cho từng lớp, chuẩn hóa các giá trị đầu ra thành các xác suất và tổng các xác suất này bằng 1.



Mô hình InceptionV3 có cấu trúc sâu và phức tạp, cho phép trích xuất các đặc trưng phức tạp từ dữ liệu hình ảnh. Với cấu trúc Inception module và các kỹ thuật tiên tiến khác, InceptionV3 đạt được kết quả ấn tượng trong việc phân loại và nhận dạng hình ảnh trên nhiều tập dữ liệu.

**2.3.2.4 ResNet50**

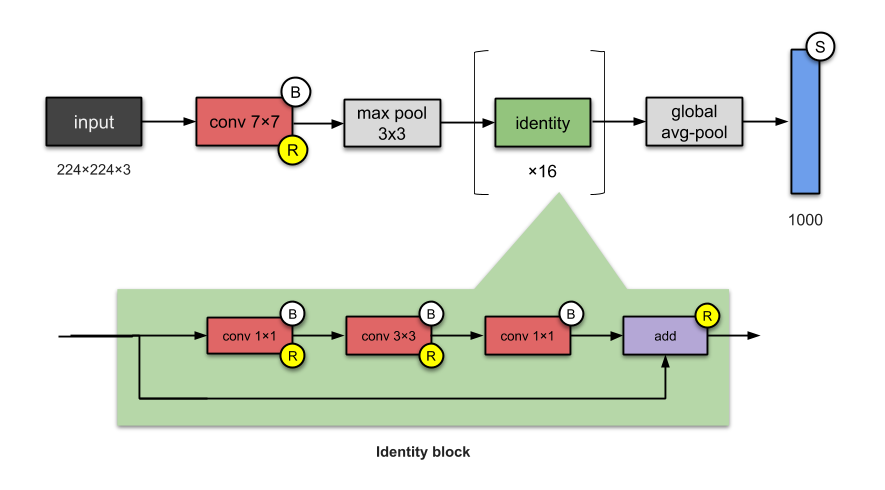
Mô hình học máy ResNet50 là một kiến trúc mạng nơ-ron sâu (deep neural network) được sử dụng phổ biến trong lĩnh vực thị giác máy tính và nhận dạng hình ảnh. Được giới thiệu bởi Kaiming He, Xiangyu Zhang, Shaoqing Ren và Jian Sun vào năm 2015, ResNet50 là một biến thể của mô hình ResNet (Residual Network) với tổng cộng 50 lớp.

Cấu trúc của ResNet50 dựa trên mô hình mạng nơ-ron tích chập (convolutional neural network - CNN) và sử dụng một kỹ thuật được gọi là "skip connection" (kết nối bỏ qua) để giúp cải thiện khả năng học sâu và tránh vấn đề vanishing gradient.

ResNet là kiến trúc sớm nhất áp dụng batch normalization. Mặc dù là một mạng rất sâu khi có số lượng layer lên tới 152 nhưng nhờ áp dụng những kỹ thuật đặc biệt mà ta sẽ tìm hiểu bên dưới nên kích thước của ResNet50 chỉ khoảng 26 triệu tham số. Kiến trúc với ít tham số nhưng hiệu quả của ResNet đã mang lại chiến thắng trong cuộc thi ImageNet năm 2015.

Những kiến trúc trước đây thường cải tiến độ chính xác nhờ gia tăng chiều sâu của mạng CNN. Nhưng thực nghiệm cho thấy đến một ngưỡng độ sâu nào đó thì độ chính xác của mô hình sẽ bão hòa và thậm chí phản tác dụng và làm cho mô hình kém chính xác hơn. Khi đi qua quá nhiều tầng độ sâu có thể làm thông tin gốc bị mất đi thì các nhà nghiên cứu của Microsoft đã giải quyết vấn đề này trên ResNet bằng cách sử dụng kết nối tắt.

ResNet có khối tích chập (Convolutional Bock, chính là Conv block trong hình) sử dụng bộ lọc kích thước 3 x 3 giống với của InceptionNet. Khối tích chập bao gồm 2 nhánh tích chập trong đó một nhánh áp dụng tích chập 1 x 1 trước khi cộng trực tiếp vào nhánh còn lại.Khối xác định (Identity block) thì không áp dụng tích chập 1 x 1 mà cộng trực tiếp giá trị của nhánh đó vào nhánh còn lại.



Tổng cộng, ResNet50 có 50 lớp (bao gồm cả các lớp convolutional, fully connected và các lớp activation) và khoảng 23,5 triệu tham số. Kiến trúc này đã được chứng minh rằng có khả năng học hiệu quả và cho kết quả tốt trên nhiều tác vụ thị giác máy tính.

**2.3.2.5 MobileNetV2**

MobileNetV2 là một mô hình học máy được phát triển cho việc nhận dạng hình ảnh trên các thiết bị di động với tài nguyên tính toán hạn chế. Mô hình này là một phiên bản cải tiến của MobileNetV1, được giới thiệu bởi nhóm nghiên cứu Google vào năm 2018.

MobileNetV2 sử dụng kiến trúc gồm các khối cơ bản là Depthwise Separable Convolution, giúp giảm đáng kể số lượng tham số và tính toán so với các mô hình truyền thống. Đây là một yếu tố quan trọng để đảm bảo khả năng triển khai trên các thiết bị có tài nguyên hạn chế.

**Chương 3: Thực nghiệm**

**3.1. Tập dữ liệu**

Để thực hiện khóa luận này, tôi đã tiến hành quay phim tại 3 tuyến đường tại Hà Nội (đường Hồ Tùng Mậu, Xuân Thủy, Tây Sơn) vào các khung giờ (từ 7h đến 8h, từ 15h đến 16h, từ 18h đến 19h). Các video sau đó được chia thành nhiều khung hình (frame). Các khung hình này có độ phân giải là 1920x1080. Sau đó, những hình ảnh này được gán nhãn thủ công bằng cách chia vào 5 thư mục tương ứng với 5 mức của mật độ giao thông. Thông tin về bộ dữ liệu dùng trong nghiên cứu được tổng hợp theo bảng dưới đây:

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

## **3.2 Thực nghiệm**

Xử lý video thành các frame ảnh

Công tác chuẩn bị dữ liệu cho mô hình Học sâu được tiến hành như sau: bộ dữ liệu ảnh giao thông được chia theo tỉ lệ 4:1 cho việc đào tạo và kiểm tra độ chính xác mô hình. Đoạn mã dùng để chia ngẫu nhiên bộ dữ liệu nhằm phục vụ dự án được viết bằng ngôn ngữ Python.

Những khung phần mềm (framework) phổ biến cung cấp tác vụ Phân loại ảnh hiện nay có thể kể đến là PyTorch, Tensorflow,… Trong đó, Tensorflow là nội bật hơn cả bởi nó được phát triển và cập nhật liên tục bởi Google, đồng thời tài liệu của Tensorflow cũng được biên soạn rất đầy đủ, chi tiết dành cho người mới làm về AI như học sinh, sinh viên tiếp cận. Tensorflow cung cấp đa dạng các mô hình Học sâu được huấn luyện sẵn trên nhiều bộ dữ liệu lớn như COCO, Kitti, Open Images, AVA v2.1,… nên hoàn toàn phù hợp với hướng nghiên cứu của khóa luận.

### **3.2.1 Tiền xử lý**

Chuẩn bị file .csv

### **3.2.2 Huấn luyện**

Tôi đã thực hiện đào tạo mô hình

Cấu hình máy tính tôi sử dụng để thực hiện Huấn luyện mô hình:

· CPU: Intel® Core™ i7–9700K @ 3.60 GHz

· GPU: NVIDIA® GeForce™ RTX 2080Ti 11GB

**Chương 4: Kết quả và đánh giá**

**4.1. Kết quả**

**4.2. Đánh giá**

**Chương 5: Tổng kết**

Qua khóa luận này, tôi đã liên hệ được sự cần thiết của việc áp dụng mô hình Học sâu Nhận diện đối tượng vào lĩnh vực giao thông, cụ thể là đo lường mật độ giao thông tự động. Ứng dụng này của Học sâu vừa có độ chính xác cao, đáng tin cậy vừa giúp rút ngắn thời gian, công sức của các nhân viên, cán bộ công tác trong lĩnh vực này.

Bằng thực nghiệm, tôi nhận thấy mạng … đem lại độ chính xác cao nhất. Về tốc độ phân loại ảnh thì mô hình … chiếm ưu thế hơn cả. (Dẫn chứng)

Trong tương lai, thực nghiệm trong khóa luận của tôi có thể được nghiên cứu thêm và tối ưu các mạng Học sâu cho bài toán này thông qua việc điều chỉnh các tham số mô hình. Bên cạnh đó, việc phát triển một ứng dụng web phục vụ bài toán “Phân loại tự động mật độ giao thông dựa trên ảnh đầu vào” bằng khung phần mềm Flask và triển khai các mô hình Học sâu đã huấn luyện.