

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



NGUYỄN TẤN HẢI

**PHÁT HIỆN ĐIỂM ÒN TẮC GIAO THÔNG
BẰNG VIDEO**

LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI – NĂM 2020

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



NGUYỄN TẤN HẢI

**PHÁT HIỆN ĐIỂM ÒN TẮC GIAO THÔNG
BẰNG VIDEO**

**Chuyên ngành: Hệ thống thông tin
Mã số: 8.48.01.04**

LUẬN VĂN THẠC SĨ KỸ THUẬT

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN VĂN THỦY

HÀ NỘI - NĂM 2020

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kết quả nêu trong luận văn là hoàn toàn trung thực và chưa từng được công bố trong bất kỳ công trình nào khác.

Học viên

Nguyễn Tấn Hải

LỜI CẢM ƠN

Trước tiên, tôi xin gửi lời cảm ơn sâu sắc nhất đến thầy giáo TS. Nguyễn Văn Thủy - Khoa Công nghệ thông tin 1 - Học viện Công nghệ Bưu Chính Viễn thông. Người đã tận tình hướng dẫn và chỉ bảo tôi trong suốt quá trình thực hiện khóa luận đồng thời cảm ơn các thầy cô Khoa Công nghệ Thông tin I, Khoa Đào tạo Sau Đại học – Học viện Công nghệ Bưu chính Viễn thông, đặc biệt là sự quan tâm giúp đỡ của thầy Phạm Văn Cường đã truyền đạt những kiến thức bổ ích cho tôi trong suốt quá trình tôi làm luận văn.

Tôi xin chân thành cảm ơn gia đình tôi đã giúp đỡ để tôi có thời gian hoàn thành khóa luận. Cuối cùng, tôi xin chân thành cảm ơn các anh chị, bạn bè và các bạn sinh viên đã giúp đỡ tôi trong quá trình thu thập và xử lý dữ liệu.

Hà Nội, ngày 15 tháng 04 năm 2020

Học viên

Nguyễn Tấn Hải

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
MỤC LỤC.....	iii
DANH MỤC CÁC CHỮ VIẾT TẮT	v
DANH MỤC CÁC BẢNG.....	vi
DANH MỤC CÁC HÌNH.....	vii
MỞ ĐẦU.....	1
Chương 1. Tổng quan về phát hiện điểm ùn tắc giao thông	3
1.1. Ý nghĩa của việc phát hiện điểm ùn tắc giao thông.	3
1.2. Các nghiên cứu liên quan	4
1.2.1. Sử dụng video từ camera giám sát giao thông	5
1.2.2. Sử dụng phương pháp học sâu:	11
1.3. Phạm vi nghiên cứu.....	14
1.4. Kết luận chương 1	14
Chương 2. Phương pháp phát hiện điểm ùn tắc giao thông	15
2.1. Mô hình mạng nơron tích chập (CNN)	15
2.1.1. Khái niệm về mạng nơron tích chập	15
2.1.2. Kiến trúc mạng nơron tích chập.....	15
2.2. Mô hình mạng SSD (Single Shot Multibox Detector)	20
2.2.1. Ưu điểm mô hình SSD	20
2.2.2. Kiến trúc mô hình SSD	21
2.3. Mô hình MobileNet.....	25

2.3.1. Giới thiệu mô hình MobileNet [16].	25
2.3.2. Kiến trúc mô hình MobileNet.	25
2.4. Phương pháp đếm phương tiện giao thông	30
2.4.1. Dữ liệu video số	30
2.4.2. Phát hiện và phân loại phương tiện trong video	32
2.4.3. Đếm mật độ phương tiện giao thông trong một khu vực.	37
2.5. Phát hiện điểm ùn tắc giao thông.	38
2.6. Kết luận chương 2	39
Chương 3. Thử nghiệm phát hiện điểm ùn tắc giao thông.	40
3.1. Dữ liệu	40
3.2. Thử nghiệm	43
3.2.1. Cài đặt môi trường thử nghiệm	43
3.2.2. Mô tả hoạt động của ứng dụng.	43
3.2.3. Kết quả thử nghiệm trên tập dữ liệu thu thập	44
3.3. Kết luận chương 3	46
Kết luận	48
DANH MỤC CÁC TÀI LIỆU THAM KHẢO.	49

DANH MỤC CÁC CHỮ VIẾT TẮT

Viết tắt	Tiếng Anh	Tiếng Việt
CNNs	Convolutional Neural Network	Mạng nơron tích chập
DCNN	Deep Convolution Neural Networks	Mạng nơron tích chập sâu
MB	Model background	Mô hình nền
MOD	Moving Object Detector	Xác định đối tượng di chuyển
MOD	Moving Object Detector	Xác định đối tượng di chuyển
NNs	Neural NetWork	Mạng nơron
ReLU	Rectified Linear Unit	Tính chỉnh đơn vị tuyến tính
TDC	Traffic density calculator	Tính toán mật độ giao thông
VI	Vehicle identifier	Mật độ phương tiện

DANH MỤC CÁC BẢNG

Bảng 1.1. – Bảng kết quả thực hiện phân lớp [13].	11
Bảng 1.2. Phân lớp FHWA theo mô hình DCNNs [8].....	12
Bảng 2.1. Mô hình kiến trúc mạng MobileNet.	26
Bảng 2.2. Bảng so sánh kết quả thực hiện tích chập phân tách và tích chập tiêu chuẩn trong mô hình MobileNet [16].	29
Bảng 3.1: Các thư viện cần cài đặt cho ứng dụng.....	43
Bảng 3.2. Bảng kết quả thử nghiệm trên tuyến đường nguyên trải.	45
Bảng 3.3. Bảng kết quả thử nghiệm trên tuyến đường Khuất Duy Tiến.	46

DANH MỤC CÁC HÌNH

Hình 1.1. Mô hình phân lớp phương tiện giao thông tổng quát.....	6
Hình 1.2. Sơ đồ xác định đối tượng giao thông.	7
Hình 1.3. Tách màu nền để nhận diện phương tiện tham gia giao thông [13].....	8
Hình 1.4. Mô hình mạng noron [13].	9
Hình 1.5. Frame được trích xuất từ video ở thành phố Istanbul, Thổ Nhĩ Kỳ [13]. ..	10
Hình 1.6. Frame được xử lý tách nền để nhận ra đối tượng [13].....	10
Hình 1.7. Phương pháp xác định và phân lớp phương tiện giao thông.....	11
Hình 1.8. Sử dụng Selective Search tìm kiếm đối tượng trong ảnh.....	12
Hình 1.9. Thực hiện mô hình DCNNs trên điều kiện thực tế [8].....	13
Hình 2.1. Kiến trúc của mô hình mạng noron tích chập.	16
Hình 2.2. Ví dụ về một bộ lọc tích chập sử dụng trên ma trận điểm ảnh.	17
Hình 2.3. Trường hợp thêm/không thêm viền trắng vào ảnh khi tích chập.....	18
Hình 2.4. Phương thức lấy giá trị trung bình và giá trị lớn nhất.....	19
Hình 2.5. Mô hình mạng SSD trích xuất các feature map.	22
Hình 2.6. Ô vuông dự đoán đối tượng.	22
Hình 2.7. Kiến trúc multibox.	23
Hình 2.8. Áp dụng bộ lọc tích chập để tạo ra các dự đoán.	24
Hình 2.9. Mô hình MobileNet.....	25
Hình 2.10. Cấu trúc của tích chập phân tách [5].....	27
Hình 2.11. Tích chập tiêu chuẩn và tích chập sâu phân tách kết hợp BN và ReLU[16].	29
Hình 2.12. Cấu trúc phân đoạn của video [2].	30
Hình 2.13. Cấu trúc tổng quát bài toán phát hiện ùn tắc giao thông.....	32

Hình 2.14. Hình ảnh kết quả xử lý sương mù.	33
Hình 2.15. Ảnh trước và sau khi cân bằng histogram.....	35
Hình 2.16. Cấu trúc mạng SSD (Single Shot Multibox Detector).....	36
Hình 2.17. Cấu trúc kết hợp SSD – MobileNet.	37
Hình 2.18. Phát hiện và phân loại phương tiện giao thông sử dụng mạng SSD – MobileNet.....	37
Hình 2.19. Đếm mật độ phương tiện một điểm giao thông trên đường.....	38
Hình 2.20. Phát hiện điểm ùn tắc giao thông.	39
Hình 3.1. Sơ đồ tuyến đường thu thập dữ liệu.	40
Hình 3.2. Dữ liệu giao thông thu thập trên tuyến đường Trần Duy Hưng.....	41
Hình 3.3. Dữ liệu giao thông thu thập trên tuyến đường Lê Văn Lương.	41
Hình 3.4. Dữ liệu giao thông thu thập trên tuyến đường Khuất Duy Tiến.	42
Hình 3.4. Dữ liệu giao thông thu thập trên tuyến đường Nguyễn Trãi.....	42
Hình 3.5. Kiến trúc tổng quát ứng dụng phát hiện điểm ùn tắc giao thông.	44
Hình 3.6. Kết quả chạy thử nghiệm video trên tuyến đường Nguyễn Trãi.....	45
Hình 3.7. Kết quả chạy thử nghiệm trên tuyến đường Khuất Duy Tiến.....	46

MỞ ĐẦU

Tính cấp thiết của đề tài

Ùn tắc giao thông là một trong những vấn nạn chung của các đô thị trên thế giới, kể cả các đô thị có hệ thống giao thông hiện đại bậc nhất. Theo các số liệu thống kê [18], ùn tắc giao thông gây thiệt hại to lớn cho các quốc gia trên thế giới, Ở nước Mỹ, mỗi năm tổn thất kinh tế do ùn tắc giao thông gây ra hơn 160 tỷ USD tương đương với 7 tỷ giờ làm và hàng triệu lít nhiên liệu. Ở Trung Quốc, đất nước đông dân nhất trên thế giới tình trạng ùn tắc giao thông còn tồi tệ hơn, điển hình ở quốc gia này là vụ việc ùn tắc giao thông diễn ra vào tháng 8/2010 trên cao tốc Bắc Kinh – Tây Tạng, hướng đi về phía thủ đô Bắc Kinh, Trung Quốc với tổng cộng là 100km đường cao tốc ùn tắc xảy ra trong vòng 21 ngày với số lượng phương tiện tham gia giao thông lên đến 30.000 phương tiện đã gây thiệt hại hết sức to lớn đối với nền kinh tế của quốc gia này. Còn Việt Nam chúng ta hiện nay, theo đánh giá từ Viện chiến lược và phát triển giao thông vận tải tại các thành phố lớn như Hà Nội và thành phố Hồ Chí Minh, ùn tắc giao thông ngày càng gây thiệt hại nghiêm trọng gây thất kinh tế và các vấn đề môi trường cho thành phố Hà Nội ước tính mỗi năm khoảng 1-1,2 tỷ USD, còn ở thành phố Hồ Chí Minh gần 1,3 tỷ USD [18].

Vấn đề ùn tắc giao thông là một trong những vấn đề lớn gây ảnh hưởng lớn đến nền kinh tế xã hội của nước ta. Với mục đích đưa công nghệ thông tin ứng dụng vào thực tế cuộc sống, cá nhân học viên xin chọn đề tài “**Phát hiện điểm ùn tắc giao thông bằng video**” để làm luận văn, hy vọng sẽ đóng góp tích cực về mặt lý luận và thực tiễn trong công tác điều tiết và phân luồng giao thông ở Việt Nam hiện nay.

Tổng quan về nghiên cứu:

Trong những thập niên gần đây, công nghệ nhận diện đối tượng trong ảnh kỹ thuật số ngày càng phát triển, nổi trội lên là các nghiên cứu về phát hiện và phân lớp phương tiện giao thông. Ở các nghiên cứu này, các camera giám sát được sử dụng

lắp đặt trên các tuyến đường giao thông như các cao tốc, tuyến đường trục chính với lưu lượng phương tiện đông đúc,... Từ đó, xác định các tham số về giao thông đưa ra những thông tin có ích cho việc phân loại phương tiện và điều tiết giao thông như các tham số về âm thanh, mật độ và vận tốc các phương tiện giao thông theo thời gian thực (realtime).

Đối tượng và phạm vi nghiên cứu:

Trong phạm vi luận văn sẽ tập trung vào các đối tượng đó là bài toán phát hiện, phân loại và đếm mật độ phương tiện giao thông thông qua quá trình xử lý hình ảnh trong video để phát hiện các điểm ùn tắc giao thông. Bản chất của quá này là sự phân tách các video thành các khung hình và phát hiện đối tượng chuyển động trong cảnh video. Đối tượng nghiên cứu luận văn: bài toán phát hiện, theo dõi và phân loại đối tượng chuyển động.

Các video được sử dụng trong luận văn là video được thu thập trong thực tế quay được trên các tuyến đường ở thành phố Hà Nội.

Chương 1. TỔNG QUAN VỀ PHÁT HIỆN ĐIỂM ÙN TẮC GIAO THÔNG

1.1. Ý nghĩa của việc phát hiện điểm ùn tắc giao thông.

Như chúng ta đã biết, mạng lưới giao thông mang lại cho loài người kết nối về kinh tế và xã hội giữa các vùng lãnh thổ trong một quốc gia hoặc các quốc gia với nhau được xây dựng với mật độ dày đặc và đồ sộ giúp cho các quốc gia và vùng lãnh thổ phát triển kinh tế - xã hội. Theo các số liệu thống kê hàng năm tổng thiệt hại mỗi năm do ùn tắc giao thông gây ra cho nền kinh tế của các quốc gia trên thế giới là cực kỳ to lớn. Còn ở Việt Nam, tại các thành phố lớn ùn tắc giao thông gây thiệt hại cho đất nước chúng ta hàng tỷ USD mỗi năm theo báo cáo thống kê của Viện Chiến Lược và phát triển giao thông vận tải [18], tại hai thành phố lớn là TP Hà Nội và TP Hồ Chí Minh thiệt hại ước tính mỗi năm lên tới 1.2 tỷ USD cùng với đó là các vấn đề về môi trường do lượng khí thải của các phương tiện giao thông gây ra.

Theo nghiên cứu của Trung tâm giao thông đô thị và nông thôn cho thấy, Hà Nội mỗi năm tốc độ tăng trưởng phương tiện giao thông so với tốc độ tăng trưởng hạ tầng cơ sở đang có nhiều bất cập làm gia tăng gánh nặng lên cơ sở hạ tầng giao thông. Mặt khác, ý thức của người tham gia giao thông còn kém tạo ra áp lực lớn đến giao thông thủ đô làm cho giao thông của TP Hà Nội ngày càng ùn tắc hơn [20]. Do đó việc phát hiện sớm các điểm ùn tắc giao thông để lực lượng chức năng có thể nhanh chóng điều tiết và phân luồng giao thông làm giảm ùn tắc, giảm đáng kể các thiệt hại về kinh tế do ùn tắc giao thông gây ra là một nhiệm vụ hết sức quan trọng và cần thiết. Ngày nay, với sự tiến bộ của lĩnh vực công nghệ thông tin, trong đó điểm hình là sự phát triển của trí tuệ nhân tạo; các phương pháp học máy và học sâu được áp dụng để giải quyết các bài toán giao thông. Các bài toán giao thông thường trải qua các bước cơ bản là phát hiện, phân loại và đếm mật độ phương tiện giao thông từ đó đưa ra dự đoán về khả năng phát hiện các điểm ùn tắc giao thông dựa trên video là xem xét tại một thời điểm, trong vùng quan tâm, có những loại

phương tiện gì, số lượng tương ứng. Về ứng dụng, bài toán này thuộc nhóm các ứng dụng liên quan đến lĩnh vực giao thông thông minh. Yêu cầu cơ bản của bài toán là với dữ liệu đầu vào là video giao thông, yêu cầu đầu ra là phân lớp phương tiện giao thông trong vùng quan sát (ô tô, xe gắn máy và các phương tiện giao thông thô sơ khác) và dự báo tại thời điểm và vị trí đó có xảy ra ùn tắc giao thông hay không.

Như vậy, việc phát hiện sớm các điểm ùn tắc giao thông có ý nghĩa cực kỳ quan trọng, từ các thông tin báo về các điểm ùn tắc giao thông sẽ giúp cho các lực lượng quản lý giao thông để các lực lượng này nhanh chóng thực hiện điều tiết và phân luồng giao thông làm giảm bớt đi thiệt hại mà ùn tắc giao thông gây ra cho nền kinh tế của nước ta.

1.2. Các nghiên cứu liên quan

Những năm gần đây, với sự tiến bộ của khoa học công nghệ đã có rất nhiều phương pháp để giải quyết các bài toán giao thông, nhưng chủ yếu được chia thành 3 nhóm phương pháp chính: nhóm phương pháp xác định cơ bản (detection based methods), nhóm phương pháp phát hiện chuyển động (motion based methods) và nhóm phương pháp kết hợp (holistic methods).

- Nhóm phương pháp phát hiện cơ bản (detection base methods) thường sử dụng video để xác định mật độ và vị trí phương tiện giao thông và sau đó thực hiện nhiệm vụ đếm số lượng phương tiện. Nổi bật lên trong nhóm này có các nghiên cứu sau: nhóm nghiên cứu Ozkurt và Camci sử dụng phương pháp mạng nơron (neural network methods) để thực hiện bài toán đếm và phân lớp phương tiện giao thông từ video thu được [13]. Bộ lọc Kalman (Kalman filter) được sử dụng để đánh giá mật độ phương tiện giao thông. Ngoài ra nhóm nghiên cứu còn sử dụng mạng nơron tích chập nhanh (Faster RCNNs) để tính toán đánh giá mật độ giao thông [17]. Tuy nhiên, kết quả thu được còn chưa được như mong muốn đối với những video có chất lượng thấp. Tiến xa hơn nữa là nhóm nghiên cứu Adu-Gyamfi et al. sử dụng các phương pháp học sâu như mạng nơron tích chập sâu (DCNNs) để phân lớp và đếm số lượng phương tiện giao thông [4]. Gần đây, Zhang và Wu sử dụng kết hợp

cả hai phương pháp học sâu và tối ưu hóa để thực hiện đếm lưu lượng giao thông từ video chất lượng thấp.

- Nhóm phương pháp phát hiện chuyển động (motion based methods) được dùng để đánh giá lưu lượng giao thông. Các nhóm nghiên cứu như Asmaa et al. sử dụng các tham số microscopic trích xuất từ phát hiện chuyển động trong video [6]. Ngoài ra họ cũng sử dụng toàn bộ chuyển động trong video để trích xuất ra các tham số macroscopic. Tuy nhiên những phương pháp phát hiện chuyển động trên cho kết quả với độ chính xác không cao.

- Nhóm phương pháp kết hợp (holistic methods) dùng để đánh giá trạng thái giao thông trên toàn bộ ảnh bức ảnh thu được. Nhóm nghiên cứu Gonclaves et al. phân lớp video ừ tất giao thông sử dụng bộ lọc Gabor (Gabor Filters) [8]. Lempitsky và Zisserman thực hiện phép chuyển đổi tuyến tính (linear transformation) trên mỗi pixel đặc trưng để đánh giá mật độ giao thông dựa trên mỗi khung ảnh của video [11]. Tuy nhiên, phương pháp này mang lại độ chính xác thấp với những video có góc rộng.

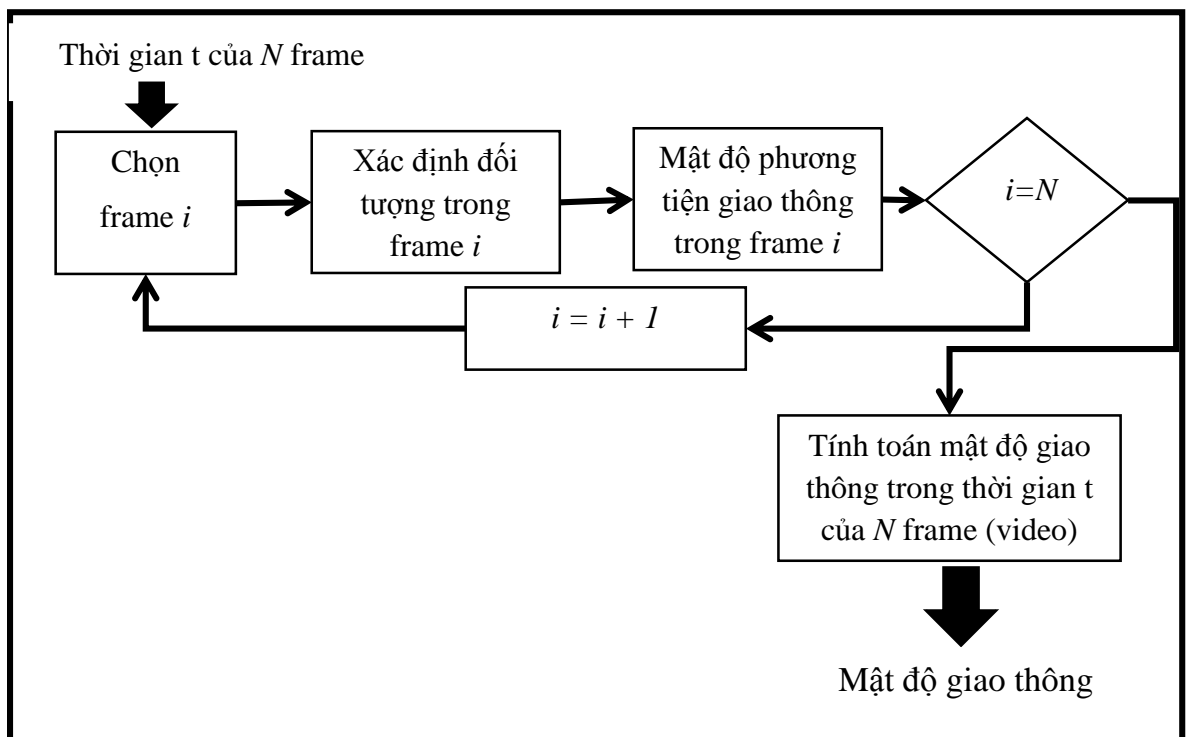
Để hiểu rõ hơn cách thức mà các nhà nghiên cứu đã sử dụng liên quan đến các nhóm nghiên cứu kể trên. Chúng ta sẽ đi qua các nghiên cứu nổi bật liên quan đến việc phát hiện và xử lý ùn tắc giao thông điển hình như sau:

1.2.1. Sử dụng video từ camera giám sát giao thông

Nhóm nghiên cứu Ozkurt và Camci [13] tiến hành nghiên cứu “Tự động đánh giá mật độ giao thông và phân lớp phương tiện sử dụng mạng nơron (NNs)” thực hiện tại thành phố Istanbul, Thổ Nhĩ Kỳ. Công ty quản lý giao thông đô thị của thành phố Istanbul, Thổ Nhĩ Kỳ đã lắp đặt và sử dụng 500 camera giám sát giao thông trên các trụ đèn giao thông với độ cao 5 mét để tăng cường giám sát và điều tiết giao thông nhằm trích xuất các thông tin hữu ích từ video thu thập được như loại phương tiện và mật độ các phương tiện tham gia giao thông nhằm phục vụ mục đích điều tiết và phân luồng giao thông.

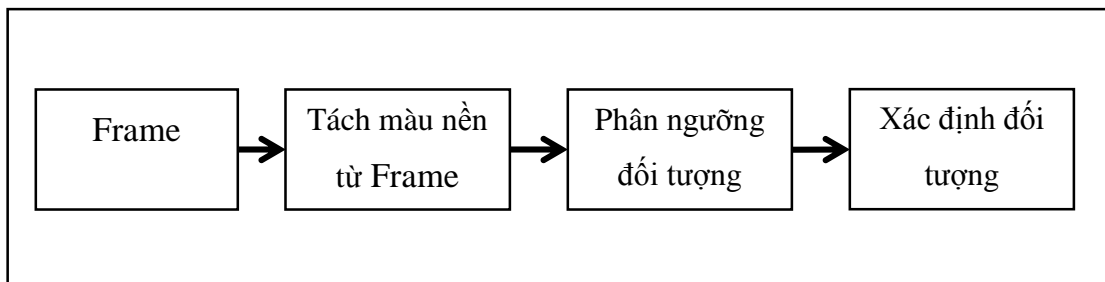
Trước đây, cũng đã có một số nghiên cứu được áp dụng thành phố Istanbul bằng cách sử dụng các cảm biến sóng, radar, hồng ngoại,...Tuy nhiên, việc thực hiện và triển khai đối với các phương pháp này tương đối phức tạp, phát sinh nhiều chi phí và khó khăn đối với các khu vực rộng lớn.

Trong nghiên cứu này các nhà khoa học đã thực hiện phân lớp phương tiện và tính toán mật độ phương tiện giao thông sử dụng mô hình nền (Model Background). Trong mô hình này, nhóm nghiên cứu thực hiện theo những bước sau đây: Đầu tiên là xác định đối tượng di chuyển và nền (Moving Object Detector - MOD) sử dụng phương pháp tách nền (background). Tiếp theo là, xác định mật độ phương tiện (Vehicle identifier – VI) thực hiện đánh giá mật độ phương tiện tham gia giao thông sử dụng model mạng nơron. Cuối cùng là tính toán mật độ giao thông (Traffic density calculator - TDC) được tính toán sử dụng từ thông tin mật độ phương tiện. Hình 1.1 dưới đây là mô hình tổng quát của hệ thống:



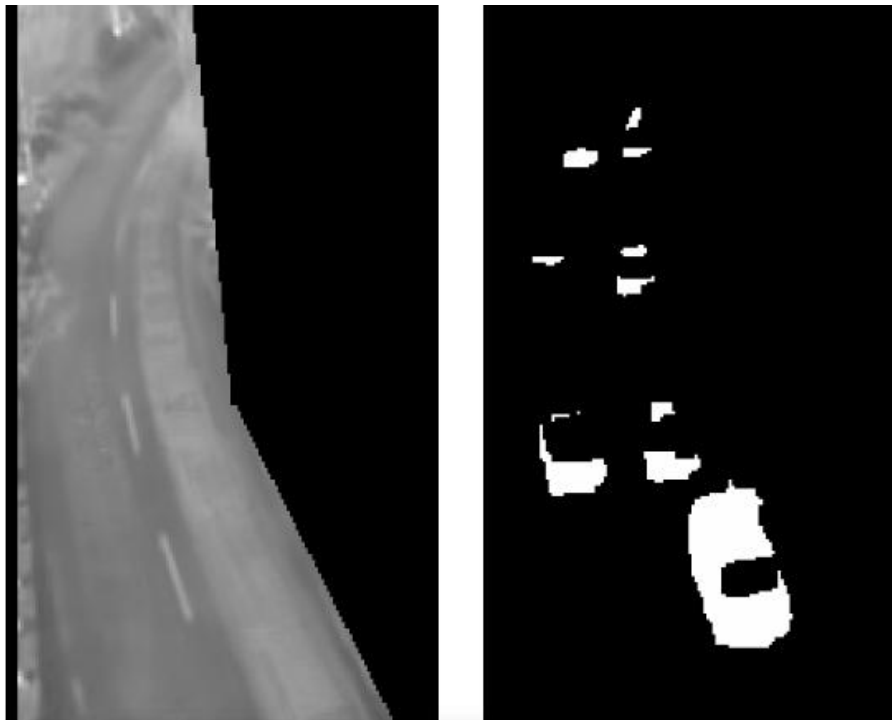
Hình 1.1. Mô hình phân lớp phương tiện giao thông tổng quát.

Bước thứ nhất là xác định đối tượng di chuyển (MOD). Các video thu thập được từ các camera quan sát. Đầu tiên, mỗi video được tách thành các khung hình (frame), sau đó trên mỗi frame của video thực hiện từng bước cụ thể theo sơ đồ hình 1.2.



Hình 1.2. Sơ đồ xác định đối tượng giao thông.

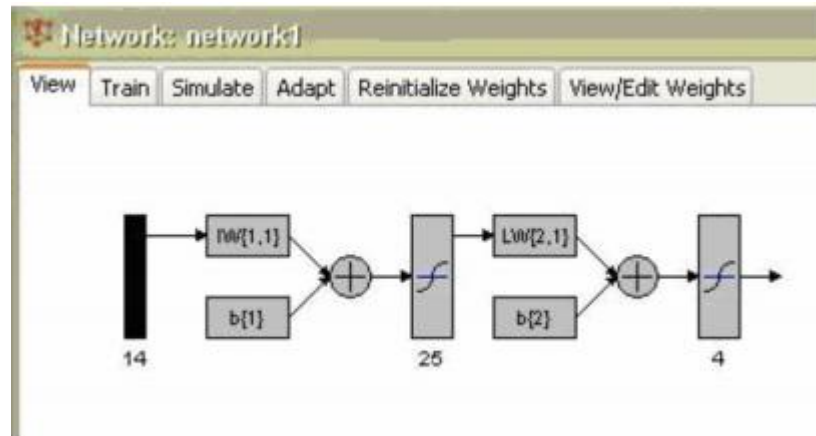
Để hiểu rõ nhóm nghiên cứu sử dụng phương pháp tách nền như thế nào để phát hiện được phương tiện giao thông và tại sao phải tách nền mới phát hiện được đối tượng tham gia giao thông. Chúng ta cùng tìm hiểu khái niệm về hình nền trong ảnh: nền hay còn được gọi là hình nền là các hình đại diện cho một quang cảnh trong video khi không có đối tượng chuyển động và phải được cập nhật thường xuyên để thích ứng với các điều kiện khác nhau về ánh sáng. Nói cách khác trong video, nền là những đối tượng đứng yên không chuyển động. Trên thực tế khó có một nền có được các đối tượng đứng yên tuyệt đối, mà luôn có những dao động nhỏ như cành lá, cây đung đưa,... Nhóm nghiên cứu đã loại bỏ nền của mỗi frame bằng phương pháp tách nền và áp dụng phân ngưỡng để chuyển sang ảnh nhị phân nhận hai giá trị 0 và 1, điểm ảnh thuộc đối tượng chuyển động được gán nhãn cho giá trị là 1 còn đối tượng không chuyển động hay là nền được gán nhãn cho giá trị là 0. Sau đó đối tượng di chuyển được xác định nhờ sự phân biệt giữa màu nền và màu đối tượng.



Hình 1.3. Tách màu nền để nhận diện phương tiện tham gia giao thông [13].

Bước thứ hai là đánh giá mật độ phương tiện giao thông. Trong phần này, nhóm nghiên cứu sử dụng mạng nơron để giải quyết bài toán phân lớp, phân cụm và đánh giá độ chính xác khi thực hiện phương pháp này. Mạng nơron được hình thành từ ý tưởng mạng lưới bộ não của con người chứa hàng triệu các liên kết nơron. Mạng nơron tính toán thông tin dữ liệu đầu ra (output) từ dữ liệu nơron đầu vào (input).

Trong bài toán này model mạng nơron được sử dụng chứa 14 layers đầu vào (input) và 4 layers đầu ra (output). Layers đầu vào chứa thuộc tính đối tượng được xác định ở phần trước. Layers đầu ra là giá trị các node nhị phân để xác định loại phương tiện (xe buýt, ô tô, xe tải) hoặc không phải là phương tiện giao thông.



Hình 1.4. Mô hình mạng nơron [13].

Cuối cùng là để tính toán mật độ giao thông (TDC) nhóm nghiên cứu sử dụng công thức 1.1 bên dưới để thực hiện tính toán số lượng phương tiện giao thông lưu thông trên đường trong một khoảng thời gian xác định. Công thức tính toán thực hiện như sau:

$$D = \frac{SL_i}{T} \quad (1.1)$$

Trong đó:

- D: mật độ giao thông (số lượng phương tiện giao thông (SL_i) / thời gian (T))
- SL_i : số lượng phương tiện giao thông di chuyển trên đoạn đường trong thời gian T .
- T : khoảng thời gian thực hiện tính toán.

Nhóm nghiên cứu áp dụng phương pháp trên vào video thực tế thu được từ camera giám sát sử dụng ở thành phố Istanbul, Thổ Nhĩ Kỳ. Trong video thu được hình ảnh giao thông từ 3 phần đường chia thành 6 làn, nhưng chủ yếu nhóm nghiên cứu tiến hành thực hiện trên phần đường nằm ở làn ngoài cùng nơi có phương tiện giao thông di chuyển với mật độ đông nhất để thực hiện áp dụng phương pháp này. Thuật toán áp dụng với 1000 khung hình trong khoảng thời gian 100 giây tức là khoảng 10 frame/s.



Hình 1.5. Frame được trích xuất từ video ở thành phố Istanbul, Thổ Nhĩ Kỳ [13].



Hình 1.6. Frame được xử lý tách nền để nhận ra đối tượng [13].

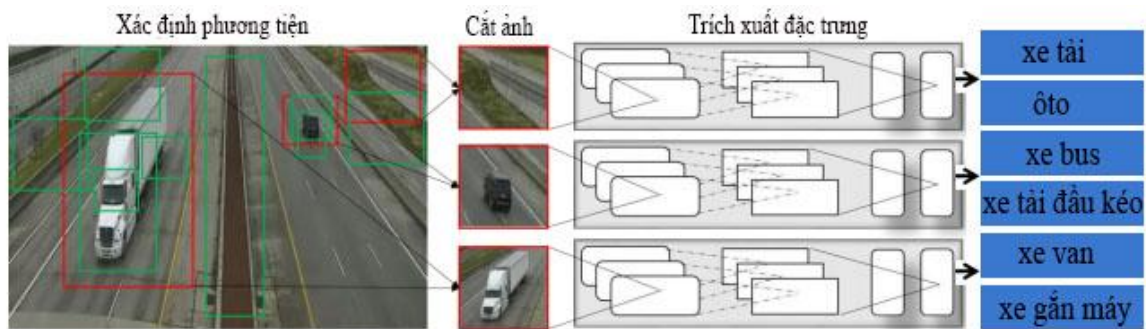
Mạng nơron xác định phương tiện giao thông đối với thời tiết bình thường tức là thời tiết bình thường với độ chính xác là 94%, phân lớp phương tiện giao thông đạt độ chính xác là 91%. Sau khi xác định và phân lớp được phương tiện giao thông nhóm nghiên cứu tiến hành tính toán mật độ giao thông. Trong 100 giây, họ đếm được có 68 phương tiện giao thông đi qua trên phần đường áp dụng phương pháp và thu được kết quả tính toán ở bảng 1.1.

Bảng 1.1. – Bảng kết quả thực hiện phân lớp [13].

	Tổng phương tiện	Xe Car	Xe Van	Xe Bus	Phân lớp
Thực tế	68	58	9	1	0
Kết quả nhận diện	68	55	8	1	4
Độ chính xác	64	64			4
Phần trăm	94%	91%			

1.2.2. Sử dụng phương pháp học sâu:

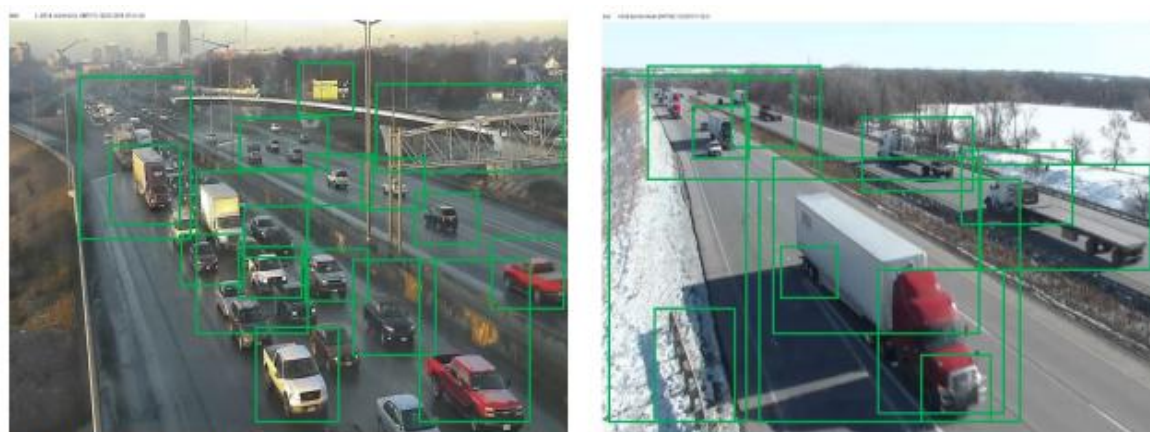
Nhóm nghiên cứu Adu-Gyamfi và Sampson Kwasi này thực hiện nghiên cứu về “ Tự động nhận diện phương tiện giao thông với mạng nơ-ron tích chập sâu (DCNNs) ” [16]. Trong nghiên cứu này, nhóm đã phát triển hệ thống camera giám sát để thực hiện hai nhiệm vụ chính là xác định vị trí và phân lớp phương tiện. Đầu tiên, nhóm nghiên cứu sẽ thực hiện xác định vị trí các phương tiện bằng thuật toán tìm kiếm (Selective Search). Sau đó, sử dụng mạng DCNNs để trích xuất các đặc trưng của phương tiện và tiến hành phân lớp.

**Hình 1.7. Phương pháp xác định và phân lớp phương tiện giao thông.**

Đầu tiên sẽ xác định vị trí của phương tiện giao thông bằng phương pháp tìm kiếm (Selective Search). Theo phương pháp truyền thống có hai cách tìm kiếm đối tượng trong ảnh là tìm kiếm phân đoạn (segmentation search) và tìm kiếm toàn phần (exhaustive search). Tìm kiếm phân đoạn là chia bức ảnh ra thành từng phần sau đó thực hiện tìm kiếm đối tượng trước khi thực hiện nhận diện đối tượng đó.

Còn đối với phương pháp tìm kiếm toàn phần thực hiện tìm kiếm toàn bộ ảnh sử dụng thuật toán cửa sổ trượt (sliding window).

Uijling et al. [16] phát triển phương pháp tìm kiếm (Selective Search) đối tượng trong ảnh, đặc điểm của phương pháp này kết hợp những ưu điểm của hai phương pháp tìm kiếm phân đoạn và toàn phần.



Hình 1.8. Sử dụng Selective Search tìm kiếm đối tượng trong ảnh.

Bước tiếp theo, tiến hành phân lớp đối tượng sử dụng mạng DCNNs. Nhóm nghiên cứu tiến hành huấn luyện mô hình DCNNs và phân lớp phương tiện theo phương pháp FHWA. Theo phương pháp phân lớp này, họ chia ra thành 13 lớp gồm có các phương tiện như sau: xe máy (motorcycle), xe oto con (car), xe khách (passenger car), xe bán tải và xe du lịch (pickup and van), xe tải, xe tải đầu kéo (single truck, multi truck).

Bảng 1.2. Phân lớp FHWA theo mô hình DCNNs [8].

Phân lớp	Tên lớp
Lớp 1	Xe gắn máy
Lớp 2	Xe ô tô con
Lớp 3	Xe tải và xe du lịch
Lớp 4	Xe buýt
Lớp 5	Xe tải
Lớp 6	Xe tải đầu kéo
Lớp 7	Xe khách

Sau đó nhóm tiến hành huấn luyện mô hình và thực hiện thử nghiệm trên thực tế với các dạng thời tiết dựa trên sự tính toán dựa trên các tham số và công thức sau: TP(đúng tích cực), FP(Sai tích cực), FN (sai tiêu cực).

- TP (True Possitive) là thực tế phương tiện giao thông trong mô hình DCNNs phát hiện đúng là phương tiện giao thông;

- FP (False Possitive) là thực tế phương tiện giao thông không xác định được khi thực hiện mô hình DCNNs;

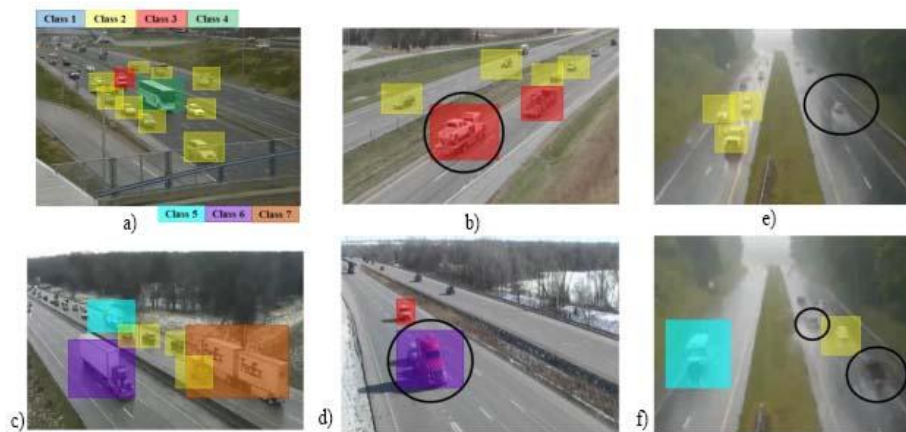
- FN (False Negative) là không xác định được đó là phương tiện giao thông;

Từ đó, chúng ta có công thức về hệ số chính xác (Precision):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1.2)$$

- Hệ số chiêu hồi (Recall):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (1.3)$$



Hình 1.9. Thực hiện mô hình DCNNs trên điều kiện thực tế [8].

Đối với các hình 1.13a) và 1.13b) trong điều kiện thời tiết có nắng nhẹ thì hệ thống hoạt động rất tốt nhận diện hầu hết các phương tiện giao thông.

Đối với các hình 1.13c) và 1.13e) trong điều kiện trời mưa một vài phân lớp phát hiện chưa chính xác.

Đối với 1.13d) và 1.13f) trong điều kiện thời tiết có tuyết rơi thì một số phương tiện không nhận diện được.

Kết quả thu được của nhóm nghiên cứu như sau về độ chính xác phát hiện phương tiện là 95% còn về độ chính xác phân lớp phương tiện là 93%.

1.3. Phạm vi nghiên cứu

Trong phạm vi luận văn sẽ sử dụng video thu thập được quay từ các tuyến đường ở thành phố Hà Nội từ đó tiến hành thực hiện phát hiện, phân lớp và xác định ùn tắc giao thông.

Luận văn chủ yếu tập trung giải quyết vấn đề phát hiện, phân loại và đếm mật độ phương tiện giao thông từ đó đưa ra dự đoán xem có ùn tắc giao thông hay không thông qua quá trình xử lý hình ảnh và đưa ra thông tin. Bản chất của quá trình bày: phân tách video thành các khung hình (frame); phát hiện đối tượng chuyển động trong video, dựa vào các đặc trưng hình học, chuyển động để phân loại và kết hợp với những đặc tính không gian, thời gian khác để đếm các loại đối tượng chuyển động có trong video.

1.4. Kết luận chương 1

Ở chương 1, luận văn đã trình bày về ý nghĩa quan trọng của việc phát hiện điểm ùn tắc giao thông, khảo sát các bài toán phát hiện và phân lớp giao thông liên quan trên thế giới đã được nghiên cứu và xác định chính xác phạm vi nghiên cứu.

Tiếp theo sang chương 2, chúng tôi sẽ trình bày về các phương pháp phát hiện điểm ùn tắc giao thông, mô tả cụ thể chi tiết các mô hình được sử dụng như CNN, SSD và MobileNet để giải quyết bài toán; phương pháp đếm và phát hiện điểm ùn tắc giao thông.

Chương 2. PHƯƠNG PHÁP PHÁT HIỆN ĐIỂM ỪN TẮC GIAO THÔNG

Chương này trình bày về các mô hình mạng nơron tích chập, mạng SSD và mô hình MobileNet áp dụng để giải quyết bài toán phát hiện, phân lớp giao thông từ đó đếm mật độ giao thông để đưa ra dự đoán về khả năng ứn tắc giao thông trên tuyến đường khảo sát. Trong phần này chủ yếu tập trung phân tích về khái niệm, kiến trúc của các mạng và mô hình được sử dụng để giải quyết bài toán giao thông.

2.1. Mô hình mạng nơron tích chập (CNN)

2.1.1. Khái niệm về mạng nơron tích chập

Mạng nơron tích chập (CNN – Convolutional Neural Network) là một mô hình mạng học sâu phổ biến nhất hiện nay, có khả năng nhận dạng và phân loại hình ảnh với độ chính xác rất cao. Hiện nay, các công ty công nghệ lớn như Google, Facebook hay Amazon đã ứng dụng mạng này vào trong các sản phẩm của họ những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe tự hành hoặc phân loại hàng hóa.

Mạng CNN được sử dụng rất nhiều trong các bài toán nhận dạng đối tượng trong ảnh. Ý tưởng của mạng CNN dựa trên sự cải tiến các mạng truyền thống. Trong khi các mạng truyền thống sử dụng các liên kết đầy đủ giữa các điểm ảnh, chính vì vậy gây ra hạn chế rất lớn bởi vì với sự phát triển công nghệ càng ngày các bức ảnh có kích thước càng lớn làm cho các liên kết càng tăng nhanh và sau đó khi thực hiện tính toán sẽ ngày càng khó khăn. Sự ra đời của mạng CNN với sự thay đổi lớn về kiến trúc có khả năng xây dựng liên kết chỉ sử dụng một phần cục bộ của bức ảnh thay vì toàn bộ bức ảnh như mạng nơron truyền thống [1].

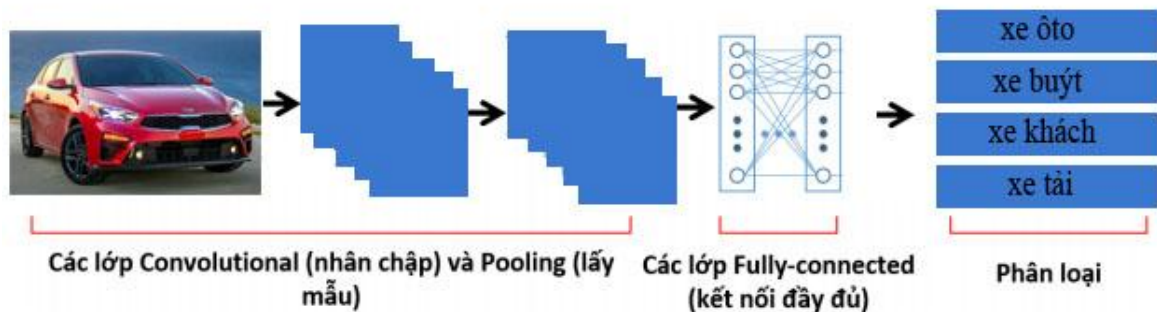
2.1.2. Kiến trúc mạng nơron tích chập

Mạng CNN là một tập hợp các lớp tích chập chồng lên nhau và sử dụng các hàm phi tuyến và tuyến tính để kích hoạt các trọng số. Mỗi một lớp thông qua các

hàm kích hoạt sẽ tạo ra các thông tin hữu ích cho các lớp tiếp theo. Về cơ bản, mạng nơron tích chập cũng như các mạng nơron truyền thống có các lớp cơ bản sau đây:

- Lớp tích chập (Convolutional layer);
- Lớp kích hoạt phi tuyến ReLU (Rectified Linear Unit);
- Lớp lấy mẫu (Pooling layer);
- Lớp kết nối đầy đủ (Fully connected layer);

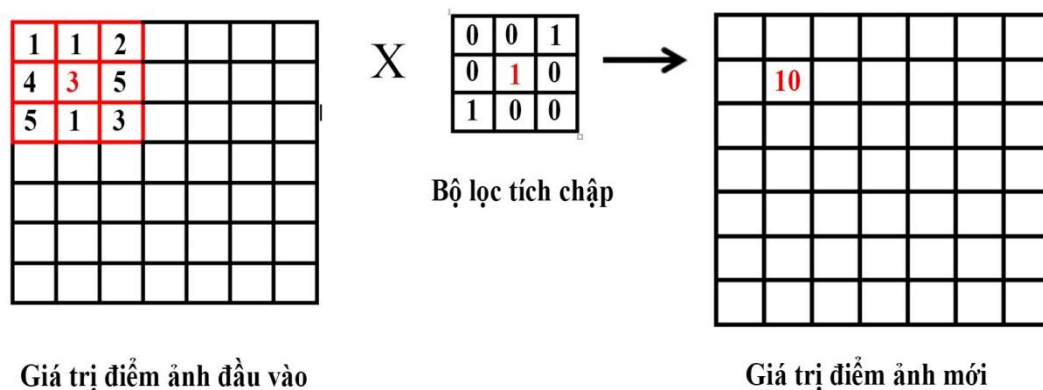
Như vậy, đối với các bài toán khác nhau thì sẽ bố trí cách sắp xếp và thay đổi số lượng 4 lớp trên khác nhau để tạo ra các mô hình huấn luyện phù hợp. Hình 2.1. mô tả kiến trúc tổng quát của mô hình mạng CNN với ảnh đầu là phương tiện giao thông vào sau đó thực hiện các lớp nhân chập, lấy mẫu. Sau khi lấy mẫu xong thì thực hiện các lớp kết nối đầy đủ và cuối cùng là phân lớp các phương tiện giao thông.



Hình 2.1. Kiến trúc của mô hình mạng nơron tích chập.

a) Lớp tích chập (Convolutional layer):

Đây là thành phần quan trọng nhất trong mạng CNN, cũng là nơi thể hiện tư tưởng xây dựng sự liên kết cục bộ thay vì kết nối toàn bộ các điểm ảnh. Lớp này có chức năng chính là phát hiện các đặc trưng. Lớp này chứa các thành phần như sau: một ma trận đầu vào, bộ lọc tích chập và các đặc trưng (feature map).



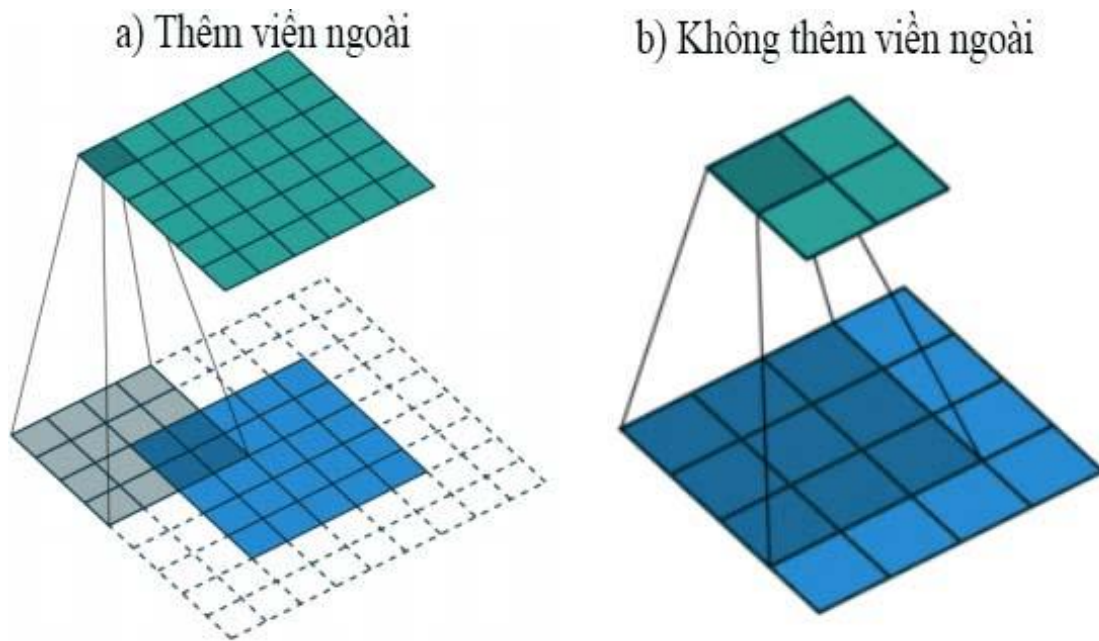
Hình 2.2. Ví dụ về một bộ lọc tích chập sử dụng trên ma trận điểm ảnh.

Trong ví dụ ở Hình 2.2, một bộ lọc tích chập được sử dụng cho phép toán nhân chập là một ma trận có kích thước 3x3 trong hầu hết các tình huống các bộ lọc tích chập đều là số lẻ. Sau đó, bộ lọc này di chuyển lần lượt qua từng vùng ảnh tức ma trận đầu vào đến khi hoàn thành quét toàn bộ bức ảnh, kết quả thu được là một bức ảnh mới có kích thước nhỏ hơn hoặc bằng kích thước ảnh đầu vào. Kích thước ảnh đầu ra tùy thuộc vào kích thước ảnh đầu vào có thêm viền vào ảnh gốc và được tính theo công thức sau (1) [7]:

$$SL = \frac{m + 2 * k - n}{st} + 1 \quad (2.1)$$

Trong đó:

- SL: kích thước ảnh đầu ra.
- m: kích thước ảnh đầu vào.
- k: kích thước khoảng trắng phía ngoài viền của ảnh gốc.
- n: kích thước bộ lọc
- st: bước trượt của bộ lọc.



Hình 2.3. Trường hợp thêm/không thêm viền trắng vào ảnh khi tích chập.

Như vậy, sau khi đưa một ảnh đầu vào cho lớp tích chập nhận được kết quả đầu ra là một loạt ảnh tương ứng với các bộ lọc đã được sử dụng để thực hiện phép tích chập. Các trọng số của các bộ lọc này được khởi tạo ngẫu nhiên trong lần đầu tiên và sẽ được cải thiện dần xuyên suốt quá trình huấn luyện.

b) Lớp kích hoạt phi tuyến (Rectified Linear Unit Layer - ReLU):

Vào năm 2010, nhà khoa học Geoffrey E. Hinton đã giới thiệu hàm ReLU (Rectified Linear Units) là hàm kích hoạt phổ biến nhất cho mạng CNN. Hàm ReLU phổ biến vì tính toán, cài đặt đơn giản và cho kết quả tốt. Hàm này thường được đặt sau lớp tích chập, hàm ReLU sẽ gán những giá trị âm thành 0 và giữ nguyên giá trị đầu vào nếu giá trị này lớn hơn 0.

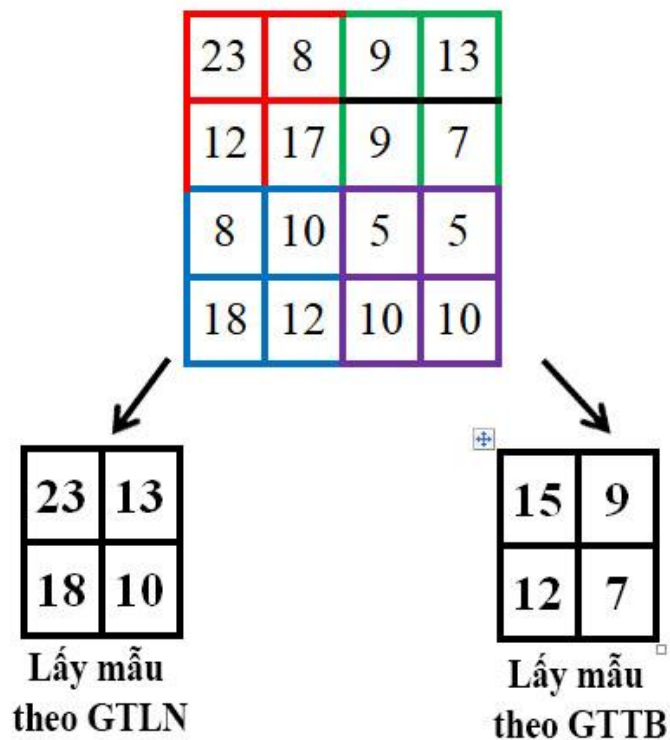
$$f(x) = \max(0, x) \quad (2.2)$$

c) Lớp lấy mẫu (Pooling layer):

Sau khi thực hiện lớp tích chập và lớp kích hoạt phi tuyến. Tiếp đến mạng CNN sẽ thực hiện một thành phần tính toán chính xác nữa là lấy mẫu (Pooling) với

mục đích làm giảm kích thước ảnh đầu ra trong khi vẫn giữ được các thông tin quan trọng của ảnh đầu vào. Tư tưởng của lớp lấy mẫu là giảm số chiều của thông tin đầu vào, hạn chế tràn bộ nhớ và giảm thời gian huấn luyện. Do vậy, lớp lấy mẫu hoạt động bằng phương pháp sử dụng một cửa sổ trượt với bộ lọc 2x2 để quét toàn bộ các vùng trong ảnh tương tự như lớp tích chập và thực hiện lấy mẫu theo giá trị lớn nhất hoặc trung bình thay vì phép tích chập tức là sẽ chọn lưu lại giá trị này đại diện cho toàn bộ thông tin của vùng ảnh đó.

Hình 2.4. thể hiện các phương thức lấy mẫu thường được dùng nhất hiện nay, đó là Max Pooling (lấy giá trị điểm ảnh lớn nhất) và Average Pooling (lấy giá trị trung bình của các điểm ảnh trong vùng ảnh cục bộ) [2].



Hình 2.4. Phương thức lấy giá trị trung bình và giá trị lớn nhất.

Như vậy, với mỗi ảnh đầu vào được đưa qua lấy mẫu thu được một ảnh đầu ra tương ứng, có kích thước giảm xuống đáng kể nhưng vẫn giữ được các đặc trưng cần thiết cho quá trình tính toán sau này.

d) Lớp kết nối đầy đủ (fully connected layer):

Lớp cuối cùng của mô hình CNN là lớp kết nối đầy đủ. Lớp kết nối đầy đủ này được thiết kế hoàn toàn tương tự như trong mạng nơron truyền thống, tức là tất cả các điểm ảnh được kết nối đầy đủ với node trong lớp tiếp theo; lớp này có chức năng chuyển các ma trận đặc trưng ở lớp lấy mẫu thành các vector chứa sắc xuất của những đối tượng cần phải dự đoán. Ví dụ: trong nhận diện phương tiện giao thông thì chuyển thành vector có 5 chiều thể hiện xác suất của 5 lớp phương tiện.

2.2. Mô hình mạng SSD (Single Shot Multibox Detector)

2.2.1. Ưu điểm mô hình SSD

Ngày nay, với sự phát triển vượt bậc của công nghệ thông tin, các mô hình mạng CNNs đang ngày càng trở nên phổ biến để giải quyết các bài toán thị giác máy tính, cùng với đó là sự ra đời mô hình RCNNs (Region Convolutional Neural Network) được phát triển dựa trên mạng CNNs nhằm giải các bài toán nhận diện và phân lớp đối tượng.

Mô hình RCNNs hoạt động theo phương pháp sau: đầu vào (input) là một ảnh, đầu ra (output) là một tập những ô vuông (bounding box) đã được gán nhãn và phân lớp đối tượng. Sau mô hình RCNNs, một vài mô hình mạng khác cũng được phát triển như: Fast RCNN, Faster RCNN nhằm cải thiện thời gian huấn luyện và độ chính xác của mô hình. Tuy nhiên các mô hình kể trên còn tồn tại một vài nhược điểm lớn như:

- Việc huấn luyện mô hình vẫn quá công kềnh và tốn nhiều thời gian.
- Quá trình huấn luyện xảy ra trên nhiều pha

- Mô hình mạng làm việc với tốc độ xử lý thấp, không đáp ứng được việc xác định đối tượng trong thời gian thực.

Những năm gần đây, những kiến trúc mạng mới được phát triển để giải quyết những các nhược điểm mà các mô hình trước kia gặp phải. Nổi bật lên là mô hình mạng SSD (Single Shot Detector). Trong mô hình này việc định vị và phân loại đối tượng được thực hiện trên 1 pha duy nhất, có khả năng nhận biết và phân loại đối tượng.

2.2.2. Kiến trúc mô hình SSD

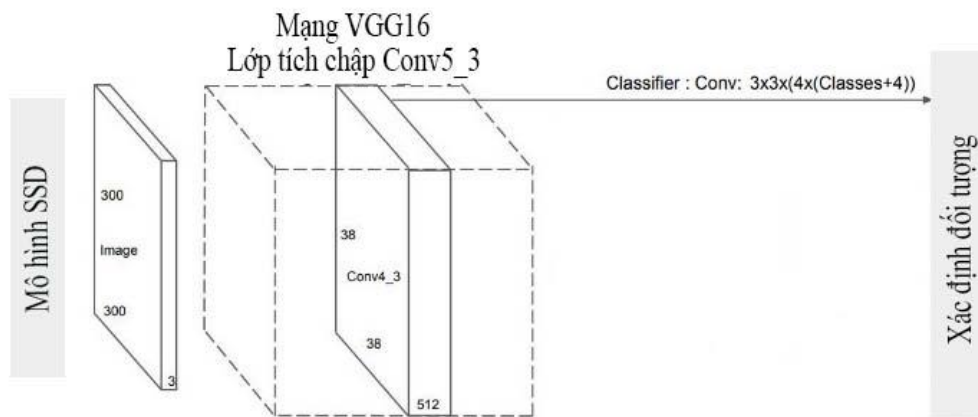
Mô hình SSD thực hiện như sau:

- Trích xuất đặc trưng (Feature Map dựa vào mạng cơ sở VGG16) để tăng hiệu quả trong việc phát hiện thì thường được sử dụng kết hợp MobileNet.
- Áp dụng các bộ lọc tích chập để có thể phát hiện các đối tượng.

a) Mô hình SSD sử dụng mạng cơ sở VGG16 để trích xuất feature map [16].

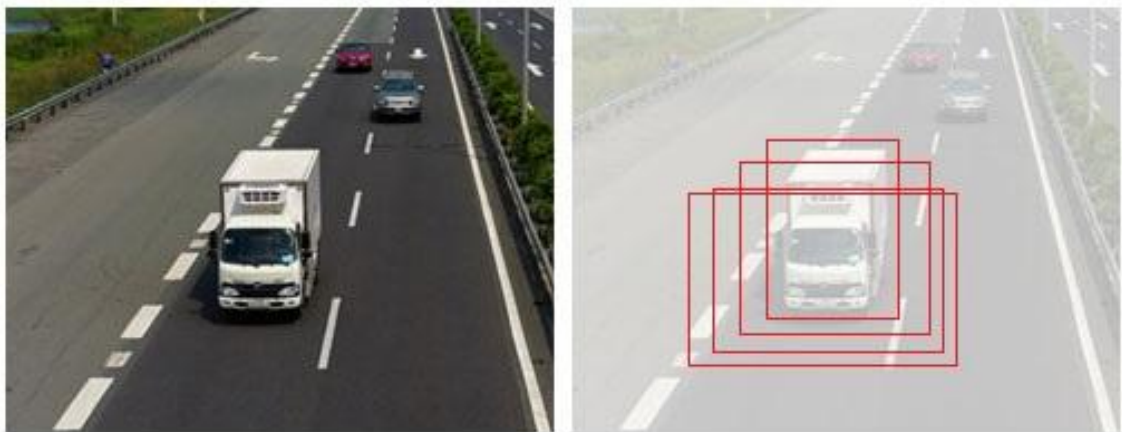
Mô hình mạng SSD sử dụng mạng VGG16 làm cơ sở mạng để trích xuất các đặc trưng (feature map) theo hình 2.5 với các tham số dưới đây:

- Input: đầu vào là ảnh có kích thước 300*300;
- Sử dụng mạng VGG16 để làm mạng cơ sở để trích xuất các đặc trưng feature map. Đối với mạng VGG16 sẽ sử dụng lớp tích chập dự đoán đối tượng (predictor object), còn các lớp còn lại chỉ có nhiệm vụ trích xuất các đặc trưng của hình ảnh đầu vào.
- Lớp dự đoán đối tượng với kích thước kernel (8,8) giảm không gian kích thước từ ảnh ban đầu là (300,300) xuống còn (38,38).



Hình 2.5. Mô hình mạng SSD trích xuất các feature map.

Sau khi tạo ra đặc trưng thì mỗi ô của từng đặc trưng sẽ đưa ra 4 ô vuông (bounding box) để dự đoán đối tượng.



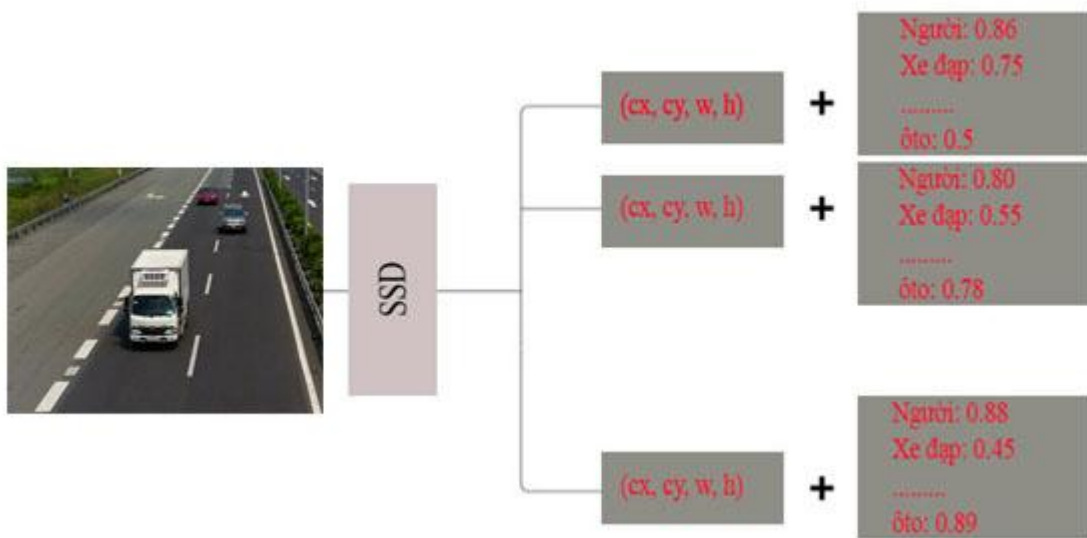
Hình 2.6. Ô vuông dự đoán đối tượng.

Hình 2.5. là hình ảnh ví dụ về ảnh gốc bên trái còn hình ảnh bên phải là hình ảnh với 1 ô là sự thể hiện của 4 bounding box dự đoán. Mỗi dự đoán bao gồm:

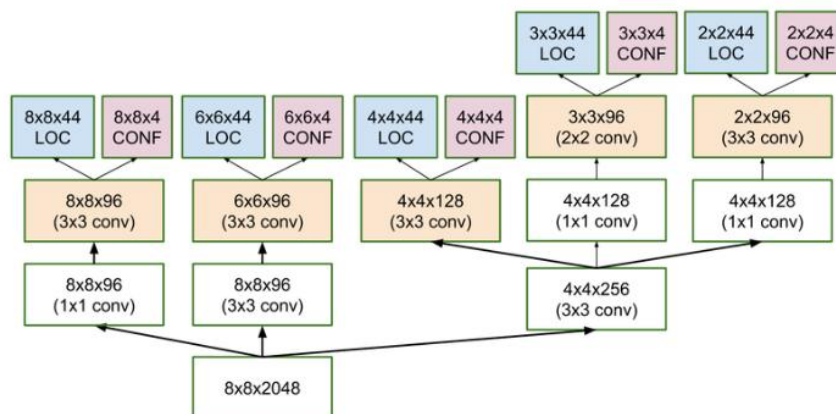
- 1 ô vuông.

- 21 score biểu thị độ tin cậy của mỗi lớp có trong ô vuông, 20 score là 20 xác suất phát hiện 20 loại đối tượng trong thực tế, 1 score còn lại là xác suất dự đoán trong ô vuông thuộc lớp nào.

Do lớp tích chập (conv4_3) đưa ra tổng số dự đoán là $38 \times 38 \times 4$ (4 bounding box được dự đoán tại mỗi vị trí pixel). Do tại mỗi pixel thì sẽ bao gồm 4 ô vuông dự đoán nên tổng số lượng ô vuông được dự đoán trên hình ảnh sẽ rất nhiều.



Kỹ thuật ô vuông dự đoán được sử dụng trong mạng SSD rất phù hợp với các đối tượng nhỏ. Hình 2.6 dưới đây là kiến trúc của một hệ nhiều ô vuông (multibox):



Hình 2.7. Kiến trúc multibox.

Hệ nhiều ô vuông dự đoán là sự kết hợp hai thành phần tương ứng với 2 chức năng của SSD:

Giá trị Cof_Loss là thành phần này giúp tính toán tỉ lệ rơi vào class mà bounding box được tính toán. Còn Giá trị loc_Loss là thành phần này ước lượng sự sai lệch thực tế của các ô vuông dự đoán so với tập dữ liệu mẫu.

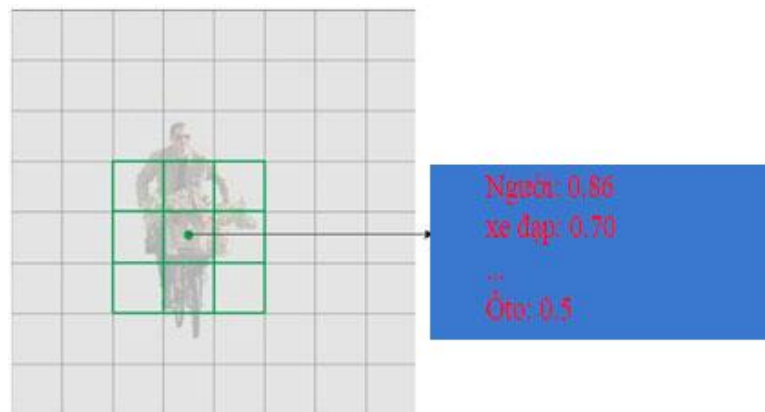
Chính vì vậy hàm mất mát (function loss) được xây dựng với 2 thành phần trên công thức như sau:

$$\text{Multibox_loss} = \text{conf_loss} + \alpha * \text{loc_loss} \quad (2.3)$$

Giá trị alpha giúp cân bằng sự ảnh hưởng của location loss. Mục tiêu tìm ra các giá trị tham số có thể tối thiểu được hàm mất mát được cho là tốt nhất, theo đó đưa ra được với dự đoán càng gần với dữ liệu mẫu.

b) Áp dụng bộ lọc tích chập cho việc phát hiện đối tượng.

Sau khi trích xuất các Feature map dựa vào kiến trúc của mạng cơ sở, mô hình SSD sẽ tiếp tục tính toán cả hai giá trị đối với từng location và score cho các lớp bằng việc sử dụng các bộ lọc tích chập. Sau khi trích xuất đặc trưng, SSD sử dụng một kernel 3*3 trên mỗi ô để tạo ra dự đoán.



Hình 2.8. Áp dụng bộ lọc tích chập để tạo ra các dự đoán.

Đối với các đặc trưng có độ phân giải thấp thì mô hình SSD thêm 6 lớp tích chập phụ trợ vào sau mạng VGG16, 5 trong số lớp đó sẽ được thêm vào để phát hiện đối tượng. Vì vậy, tổng cộng mạng SSD đưa ra 8732 dự đoán với việc sử dụng 6 lớp.

Kiến trúc mạng SSD được xây dựng trên mạng VGG16 được loại bỏ tầng kết nối đầy đủ. Mạng SSD sử dụng mạng VGG16 ở tầng cơ sở vì sự hiệu quả của nó trong bài toán phân loại ảnh với các ảnh có độ phân giải cao. Thay vì sử dụng tầng kết nối đầy đủ của VGG, một tập các tầng tích chập phụ trợ được thêm vào, vì vậy trích xuất được các đặc trưng với nhiều tỉ lệ khác nhau và giảm dần kích thước của đầu vào trong từng tầng mạng.

2.3. Mô hình MobileNet.

2.3.1. Giới thiệu mô hình MobileNet [5].

Mô hình MobileNet là một mô hình học sâu do nhóm tác giả đến từ Google tạo ra. Mô hình MobileNet được thiết kế với mục tiêu là xây dựng một mô hình mạnh mẽ nhưng nhỏ gọn, có thể thực hiện được ứng dụng trên các thiết bị di động như điện thoại, máy tính bảng hoặc các thiết bị nhúng.



Hình 2.9. Mô hình MobileNet.

2.3.2. Kiến trúc mô hình MobileNet.

Mô hình MobileNet là mô hình sử dụng cách tích tích chập đặc biệt có tên là tích chập sâu phân tách (Depwise Separable Convolution) để giảm kích thước của mô hình và giảm độ phức tạp tính toán. Do đó, mô hình này rất nhẹ và hoạt động ổn

định trên các ứng dụng di động và các thiết bị nhúng. Nhờ có cách tính tích chập sâu phân tách (Depthwise Separable Convolution) nên mô hình MobileNet có nhiều ưu điểm đó là:

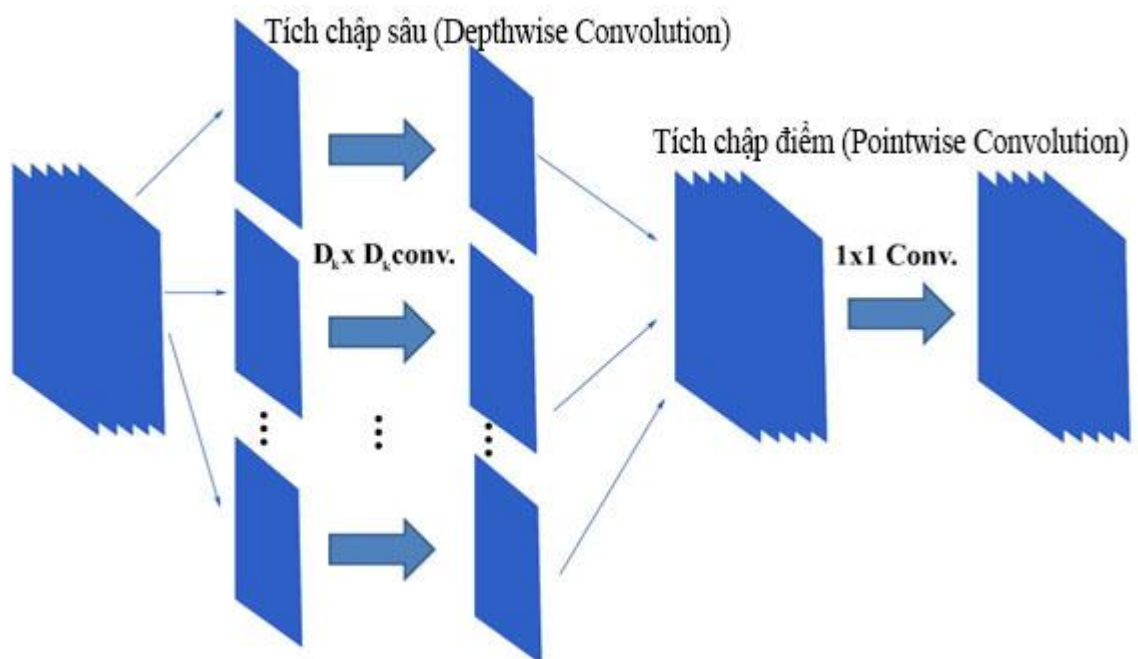
- Mô hình có ít tham số do vậy mô hình nhỏ gọn và chạy nhanh hơn.
- Số lượng các phép tính trong mô hình rất ít chính vì vậy sẽ giảm độ phức tạp.

Bảng 2.1. Mô hình kiến trúc mạng MobileNet.

	Type/ Stride (Dạng tích chập sử dụng)	Filter Shape (Bộ lọc)	Input Size (kích thước đầu vào)
1	Conv / s2	3x3x3x32	224x224x3
2	Conv dw / s1	3x3x32 dw	112x112x32
3	Conv / s1	1x1x32x64	112x112x32
4	Conv dw / s2	3x3x64 dw	112x112x64
5	Conv / s1	1x1x64x128	56x56x64
6	Conv dw / s1	3x3x128 dw	56x56x128
7	Conv / s1	1x1x128x128	56x56x128
8	Conv dw/ s2	3x3x128 dw	56x56x128
9	Conv / s1	1x1x128x256	28x28x128
10	Conv dw/ s1	3x3x256 dw	28x28x256
11	Conv / s1	1x1x256x256	28x28x256
12	Conv dw / s2	3x3x256 dw	28x28x256
13	Conv / s1	1x1x512x512	14x14x256
14	5* Conv dw/s1 Conv / s1	3x3x512dw 1x1x256x512	14x14x512 14x14x512
15	Conv dw/ s2	3x3x512 dw	14x14x512
16	Conv / s1	1x1x512x1024 dw	7x7x512
17	Conv dw/ s2	3x3x1024 dw	7x7x1024
18	Conv / s1	1x1x1024x1024	7x7x1024
19	Avg Pool/ s1	Pool 7x7	7x7x1024
19	FC/s1	1024x1000	1x1x1024
20	softmax	Classifier	1x1x1000

Trong bảng 2.1. kiến trúc mô hình MobileNet thực hiện đến 30 lớp nhưng về cơ bản mô hình này tiến hành theo 4 lớp chính và chủ yếu như sau: lớp tích chập (Convolution Layer), lớp tích chập sâu (Depthwise layer), lớp tích chập điểm (Pointwise layer), lớp softmax dùng để phân lớp phương tiện giao thông.

Trong mô hình kiến trúc MobileNet quan trọng nhất là cách tính tích chập sâu phân tách (Depthwise Separable Convolution). Tích chập sâu phân tách là một tích chập sâu theo sau bởi một tích chập điểm được thể hiện theo hình 2.9 dưới đây:



Hình 2.10. Cấu trúc của tích chập phân tách [5].

Chúng ta hãy cùng làm rõ các khái niệm về: tích chập sâu, tích chập điểm. Tích chập sâu (Depthwise convolution) là một bộ lọc áp dụng vào mỗi kênh đầu vào. Tích chập điểm (Pointwise convolution) sử dụng tích chập 1×1 để kết hợp đầu ra tích chập sâu.

Như vậy, với M là số lượng kênh đầu vào, N là số lượng kênh đầu ra, D_k là kernel size, D_f là kích thước của feature map. Khi đó chúng ta tính toán tích chập điểm và tích chập sâu dựa vào những công thức dưới đây:

Tích chập sâu được tính theo công thức sau:

$$D_k . D_k . M . D_f . D_f \quad (2.3)$$

Công thức tính tích chập điểm là:

$$M . N . D_f . D_f \quad (2.4)$$

=> Từ đó chúng ta có công thức tính tổng tích chập sâu phân tách như sau:

$$D_k . D_k . M . D_f . D_f + M . N . D_f . D_f \quad (2.5)$$

Nếu không sử dụng tích chập sâu phân tách mà sử dụng tích chập tiêu chuẩn (standard convolution) thì được tính như sau:

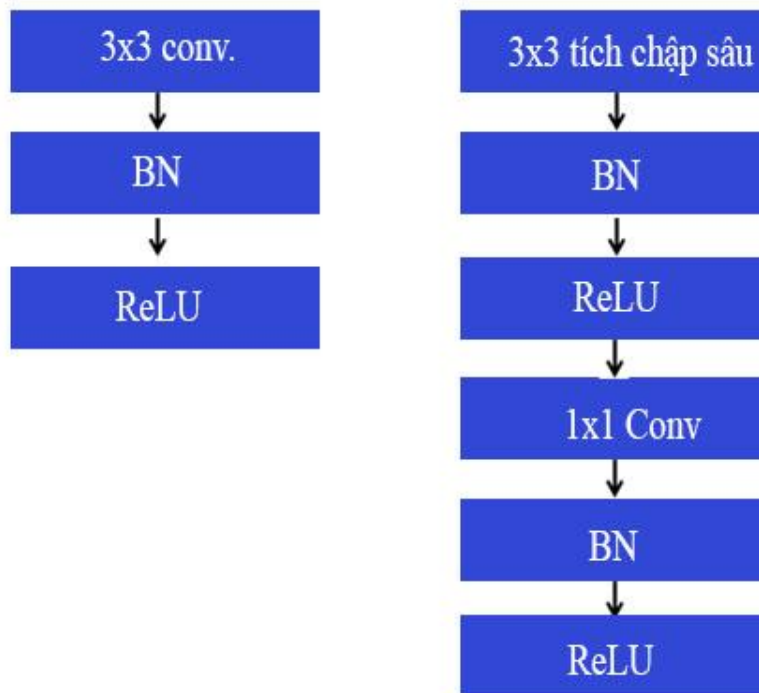
$$D_k . D_k . M . N . D_f . D_f \quad (2.6)$$

Khi đó thực hiện tính toán trên tích chập phân tách sẽ giảm đi rất nhiều so với cách tính tích chập tiêu chuẩn:

$$\frac{D_k . D_k . M . D_f . D_f + M . N . D_f . D_f}{D_k . D_k . M . N . D_f . D_f} = \frac{1}{N} + \frac{1}{D_k^2} \quad (2.7)$$

Ví dụ với kernel size $D_k = 3$ trong cách tích chập phân tách sẽ giảm tới 9 lần phép tính nhân. Chính vì vậy, giảm số lượng tính toán đi rất nhiều đó là ưu điểm rất lớn của mô hình MobileNet.

Mặt khác, sau mỗi lần tính toán tích chập mô hình MobileNet sẽ sử dụng Batch Normalization (BN) và ReLU như hình 2.10 bên dưới:



Hình 2.11. Tích chập tiêu chuẩn và tích chập sâu phân tách kết hợp BN và ReLU[16].

So sánh kết quả việc sử dụng mạng 30 layers sử dụng thuần convolution và mạng 30 layers sử dụng tích chập sâu phân tách trên tập dữ liệu ImageNet thu được kết quả sau:

Bảng 2.2. Bảng so sánh kết quả thực hiện tích chập phân tách và tích chập tiêu chuẩn trong mô hình MobileNet [16].

Model	Độ chính xác	Số lượng phép tính	Tham số
MobileNet sử dụng tích chập tiêu chuẩn	71.7%	4866	29.3
MobileNet sử dụng tích chập phân tách	70.6%	560	4.2

Tuy mô hình MobileNet giảm 1% độ chính xác, nhưng số lượng tham số của mô hình và số lượng phép tính toán giảm đi rất nhiều gần xấp xỉ 90%. Đó là một trong ưu điểm rất lớn của mô hình MobileNet.

2.4. Phương pháp đếm phương tiện giao thông

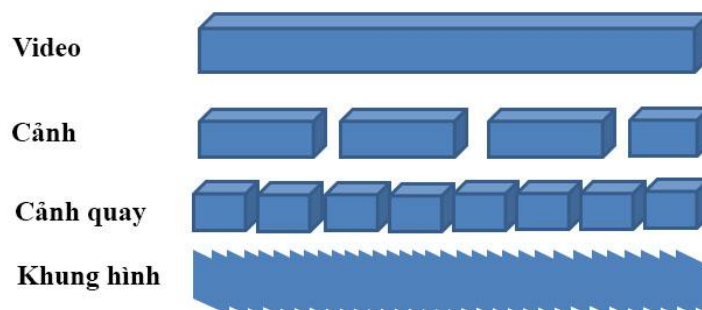
2.4.1. Dữ liệu video kỹ thuật số

Công nghệ video xuất hiện lần đầu tiên và được phát triển cho các hệ thống truyền hình cơ học, sau đó đã được thay thế bằng hệ thống truyền hình ống tia âm cực (CRT). Đến năm 1951, máy ghi băng video đầu tiên ghi lại hình ảnh trực tiếp từ máy quay truyền hình bằng cách chuyển đổi các xung điện của máy ảnh và lưu thông tin vào băng video từ tính. Sau đó năm 1971, hãng công nghệ Sony bắt đầu đưa ra thị trường các máy ghi âm và băng ghi hình video đầu tiên. Việc sử dụng các kỹ thuật số trong video đã tạo ra video kỹ thuật số, cho phép chất lượng video cao hơn và cuối cùng, chi phí thấp hơn nhiều so với công nghệ analog trước đó. Cho đến ngày nay, các video kỹ thuật số ngày càng có độ phân giải cao và độ sâu màu ngày càng cải thiện [21].

a) Khái niệm video kỹ thuật số

Video kỹ thuật số là video gồm một dãy các khung hình (frame) liên tiếp, mỗi khung hình tương ứng với một hình ảnh. Khi video thực hiện, các hình ảnh sẽ lần lượt chạy trong một thời gian nhất định, thông thường tốc độ sẽ được đặt trong khoảng 24,25 hoặc 30 khung hình mỗi giây [2].

Hình 2.9 thể hiện cấu trúc tổng quát dữ liệu của video: khung hình, cảnh quay, cảnh.



Hình 2.12. Cấu trúc phân đoạn của video [2].

Trong kỹ thuật xử lý dữ liệu video là cách sử dụng các phương pháp tác động đến nhiều thành phần của gồm có: ảnh, điểm ảnh, đường viền, nền, phép cộng ảnh, phép nhân ảnh với 1 số, ngưỡng, phân ngưỡng...

b) Đặc trưng của video kỹ thuật số

Đặc trưng của video kỹ thuật số bao gồm: màu sắc, độ phân giải, cấu trúc, hình dạng.

- Màu sắc (color): màu sắc là một đặc trưng cơ bản của ảnh. Biểu đồ màu không phụ thuộc vào việc quay ảnh, dịch chuyển ảnh, hướng ảnh mà phụ thuộc vào hệ màu. Trong video kỹ thuật số chủ yếu sử dụng hệ màu RGB (Red – Blue - Green), mọi màu sắc của chuẩn màu này được tạo từ 3 màu chính là đỏ - xanh da trời – xanh lá cây. Máy ảnh và các loại màn hình sử dụng cơ chế hòa trộn màu RGB để ghi nhận và tái tạo màu sắc. Màu RGB chỉ có thể thực hiện trên các vật có khả năng phát ra ánh sáng.

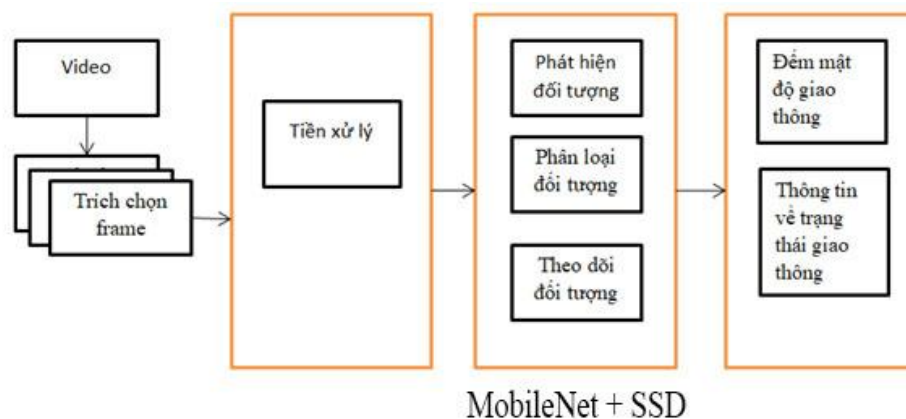
- Độ phân giải video là số lượng các điểm ảnh được hiển thị trong các thiết bị kỹ thuật số và thường được thể hiện dưới dạng chiều dài x chiều rộng (pixel). Độ phân giải video chia thành các loại sau: dạng chuẩn SD, HD, FHD, ... Độ phân giải SD (Standard Definition) có nghĩa là hình ảnh có độ nét chuẩn. Chuẩn SD thường sử dụng các độ phân giải 640x360, 640x480,... chính vì độ phân giải thấp nên các video SD thường có dung lượng rất nhỏ. Video HD thường có độ phân giải 1280x720 với độ phân giải này thì mật độ điểm ảnh lớn hơn nhiều nên các video này có độ sắc nét và chi tiết cao.

- Cấu trúc (Texture) của video được biểu diễn hai dạng phổ biến là: ma trận đồng thời và tamura. Ma trận đồng thời mô tả hướng và khoảng cách giữa các điểm ảnh, ta có thể chọn được các thống kê có ý nghĩa. Biểu diễn Tamura bao gồm các thuộc tính đo tính thô, độ tương phản, hướng, tính trơn, tính cân đối và độ thô ráp. Các đặc tính này rất quan trọng trong việc tìm hiểu nội dung ảnh vì nó biểu diễn rất trực quan [2].

- Hình dạng (Shape) đặc trưng hình dạng có thể được phân chia thành đặc trưng toàn cục và đặc trưng cục bộ. Đặc trưng toàn cục là đặc trưng thu được từ toàn bộ hình dáng đối tượng trong ảnh (ví dụ: chu vi, tính tròn, hướng trục chính). Đặc trưng cục bộ là đặc trưng thu được từ việc thao tác với một phần của ảnh, không phụ thuộc vào toàn bộ ảnh [2].

2.4.2. Phát hiện và phân loại phương tiện trong video

Bài toán phát hiện đối tượng chuyển động đã được các nhà khoa học trên thế giới nghiên cứu từ rất sớm. Cho tới hiện nay, đã có nhiều thuật toán phát hiện chuyển động được công bố. Một số thuật toán được công bố trong những năm qua đã chứng minh có độ chính xác tương đối cao, thời gian tính toán thấp. Bài toán phân loại và phát hiện đối tượng thực hiện theo cấu trúc như sau:



Hình 2.13. Cấu trúc tổng quát bài toán phát hiện ùn tắc giao thông.

Sau khi thu thập video chúng ta sẽ tiến hành:

a) Trích chọn khung hình (frame):

Đặc trưng cơ bản của video là cấu thành và sự kết hợp của các khung hình liên tiếp. Do vậy, từ dữ liệu video (trực tiếp từ máy quay camera IP hoặc từ file video), để lấy ảnh vào phân tích, hệ thống phải tách thành các khung hình (chứa ảnh đối tượng) để tiến hành trừ nền, phát hiện đối tượng,...

b) Tiền xử lý ảnh

Trong quá trình thu thập video sử dụng các thiết bị camera, điện thoại, UAV,... Các thiết bị này tuy hiện tại độ phân giải tương đối tốt nhưng vì nhiều nguyên nhân như điều kiện sáng, thời tiết không tốt gây ra cho video thu thập bị kém chất lượng. Vì vậy, để tăng hiệu quả cho quá trình phát hiện và phân loại phương tiện giao thông cần phải tiền xử lý ảnh để đạt được hiệu quả cao nhất. Trong bước này, chúng ta sẽ thực hiện khôi phục và tăng cường ảnh.

Đầu tiên sẽ thực hiện khử sương mù:

Trong điều kiện thời tiết không tốt như nắng, mưa, sương mù,... video sau khi được thu thập sẽ bị nhiễu nhưng một trong những vấn đề lớn nhất đối với bài toán giao thông gây ra nhiễu cho ảnh đó là sương mù, việc khử sương mù là một trong những công việc được đặt lên hàng đầu. Một phương pháp khử sương mù đơn giản đó là dò tìm các điểm tối, đây là phương pháp hiệu quả để loại bỏ sương mù trong ảnh có sương mù. Điểm tối là điểm ảnh ở trong ảnh không bị ảnh hưởng bởi sương mù với giá trị trong tất cả các kênh màu RGB gần với giá trị không. Với phương pháp này, thuật toán đề xuất có thể ước lượng một cách có hiệu quả độ dày của sương mù, lọc và tìm lại một cảnh không bị ảnh hưởng bởi sương mù, thuật toán này thậm chí còn làm việc hiệu quả với khung cảnh rộng. Ngoài ra, chất lượng của ảnh sau khi được khử sương mù rất tốt.



a)



b)

Hình 2.14. Hình ảnh kết quả xử lý sương mù.

Trong hình 2.14a) là hình mà phương tiện giao thông trên đường bị sương mù che khuất, sau khi thực hiện phương pháp khử sương mù thì ta thu được hình 2.14b) làm cho chất lượng hình ảnh trở nên tốt và dễ dàng nhận biết được các phương tiện giao thông.

Tiếp theo sẽ thực hiện phương pháp tăng cường ảnh

Sau khi khử sương mù, vấn đề tiếp theo cần thực hiện đó là nâng cao chất lượng ảnh. Trong phạm vi luận văn, sử dụng phương pháp tăng cường ảnh dựa trên cân bằng histogram. Tăng cường ảnh là phương pháp làm nổi bật các đặc trưng trong ảnh không phải là làm gia tăng thêm thông tin có trong ảnh.

Phương pháp cân bằng Histogram là sự điều chỉnh histogram về trạng thái cân bằng, giá trị các điểm ảnh không bị co cụm tại một khoảng nhỏ mà được “kéo dẫn” ra tức là tìm một phép biến đổi phi tuyến $g = T(f)$ áp dụng cho mỗi pixel của ảnh $f(x,y)$, để cho $g(x,y)$ phân bố đều [2].

Phương pháp cân bằng histogram thực hiện qua các bước như sau:

Bước 1: Thống kê số lượng pixel i trong hình ảnh.

$$p_x(i) = n_i \quad 0 \leq i < L \quad (2.8)$$

Trong đó L là số mức xám trong hình ảnh (thường là 256)

Bước 2: Tính hàm phân phối tích lũy $cdf(i)$

$$cdf_x = \sum_{j=0}^i p_x(j) \quad (2.9)$$

Bước 3: Tính lại mức xám của ảnh y sau khi cân bằng histogram từ ảnh x .

$$y(i) = \text{round}\left(\frac{cdf(i) - cdf_{\min}}{M * N - 1} * (L - 1)\right) \quad (2.10)$$

Trong đó $M*N$ là số điểm ảnh của ảnh.



a) Ảnh trước



b) Ảnh sau

Hình 2.15. Ảnh trước và sau khi cân bằng histogram.

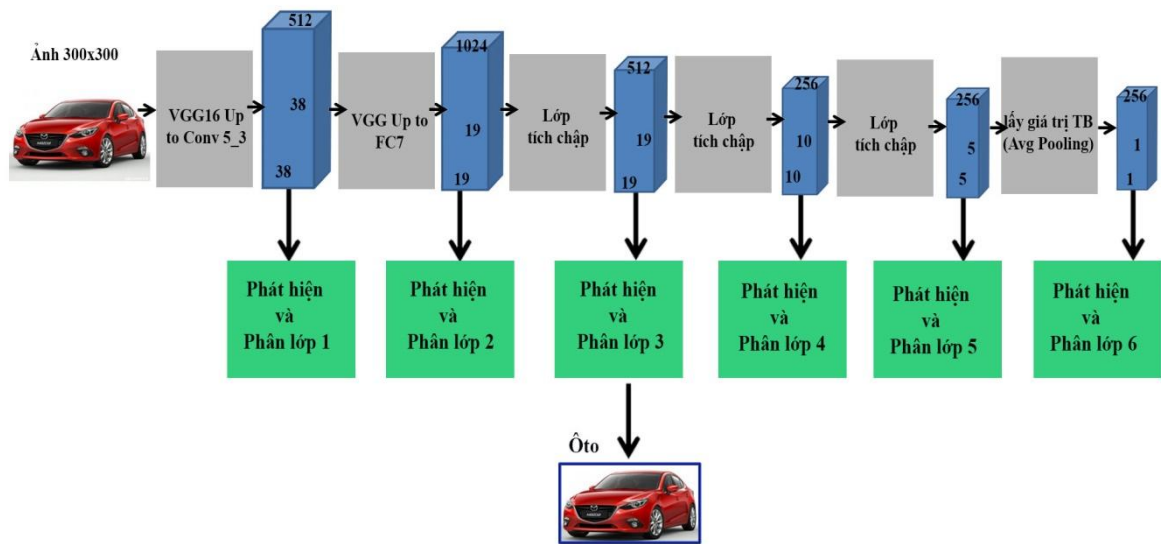
Hình 2.5 thực hiện phương pháp histogram để tăng cường ảnh làm nổi bật biên của các phương tiện giao thông với hình 2.5a) là ảnh trước khi thực hiện cân bằng histogram và hình 2.5b) là ảnh sau khi cân bằng histogram.

c) Phát hiện và phân loại phương tiện giao thông

Phát hiện đối tượng nói chung và phương tiện giao thông nói riêng là một trong những thách thức đối với các thuật toán thị giác máy tính, vì nó liên quan nhiều đến sự kết hợp giữa phân loại đối tượng và định vị đối tượng trong khung hình. Với sự tiến bộ vượt bậc của ngành học sâu, với phần lớn các nghiên cứu tập trung vào việc thiết kế các mạng phát hiện đối tượng ngày càng phức tạp hơn để cải thiện độ chính xác như SSD, Faster R-CNN, CNNs, YOLO. Hiệu suất phát hiện đối tượng trên ImageNet và PASCAL VOC đã được cải thiện đáng kể với sự phát triển của các giải thuật xử lý ảnh dựa trên mạng nơon tích chập (CNNs). Tuy vậy các giải thuật đều đòi hỏi lượng tài nguyên trên hệ thống lớn và thời gian xử lý còn chưa nhanh.

Từ những yêu cầu đó trong luận văn của cá nhân tôi đã áp dụng cấu trúc mô hình MobileNet để thực hiện phát hiện và phân lớp phương tiện giao thông bởi vì mô hình này được thiết kế nhỏ gọn, có thể thực hiện được ứng dụng trên các thiết bị

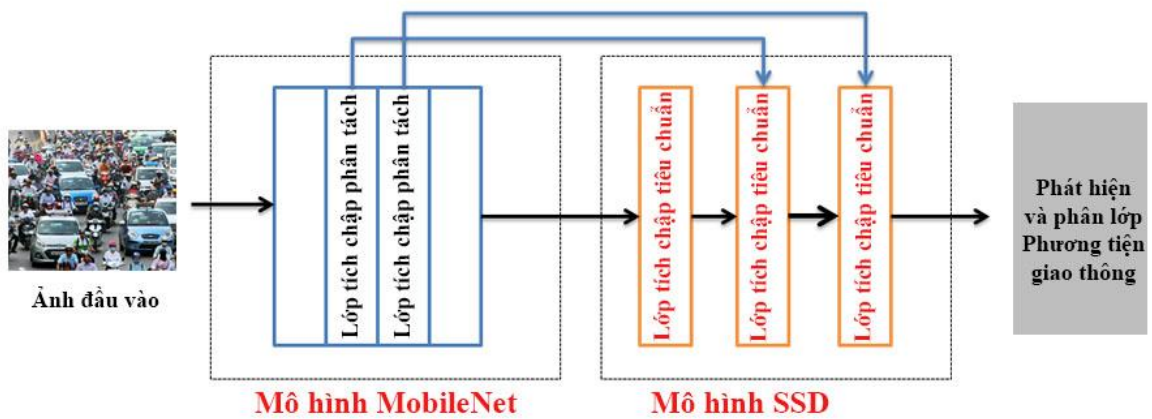
di động như điện thoại, máy tính bảng hoặc các thiết bị nhúng và thực hiện tính toán trên thời gian thực (realtime) nhằm cải tiến tốc độ và độ chính xác. Hình 2.13 dưới đây là cấu trúc tổng quát của mạng SSD



Hình 2.16. Cấu trúc mạng SSD (Single Shot Multibox Detector)

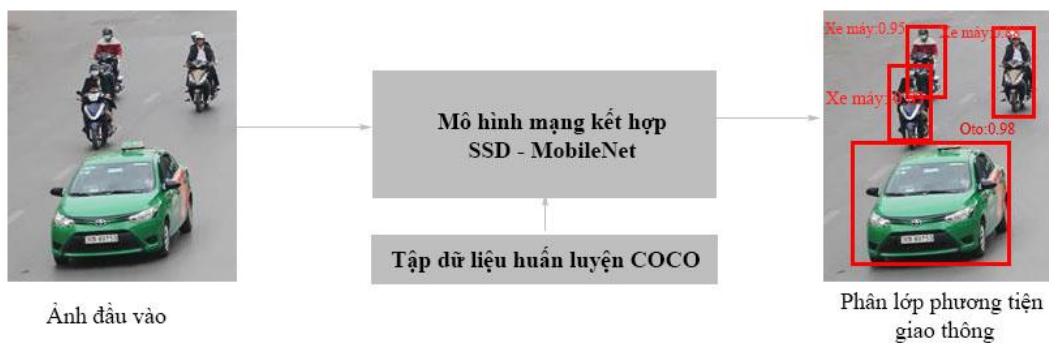
Trong hình 2.16 ảnh đầu thông qua các lớp tích chập xác định được các đặc trưng sau đó thực hiện phát hiện và phân lớp phương tiện theo nhiều bước. Rồi từ đó dựa vào xác suất có tỉ lệ cao nhất để dự đoán phương tiện giao thông.

Trong cấu trúc MobileNet chứa lớp tích chập hoàn toàn ban đầu với 32 bộ lọc, sau đó là 19 lớp bottleneck. Ngoài ra, để tăng độ chính xác của mô hình cá nhân học viện còn thực hiện kết hợp mạng SSD và MobileNet để thay thế lớp tích chập thông thường (regular convolutions) bằng các lớp tích chập phân tách (separable convolutions) trong các lớp dự đoán mạng SSD nhằm giúp mô hình giảm số lượng tham số so với các mạng tích chập thông thường, ngoài ra giúp mô hình giảm được một số bước tính toán. Cấu trúc kết hợp SSD – MobileNet được biểu diễn như hình 2.14.



Hình 2.17. Cấu trúc kết hợp SSD – MobileNet.

Các thử nghiệm của tác giả M. Sandler [15] chứng minh kết quả kết hợp mạng SSD – MobileNet bằng cách thay thế tích chập phân tách cho tích chập tiêu chuẩn trong mạng SSD mang lại hiệu quả cao hơn 20 lần và dung lượng nhỏ hơn 10 lần so với phương pháp sử dụng YOLO cũng cùng trên tập dữ liệu COCO.

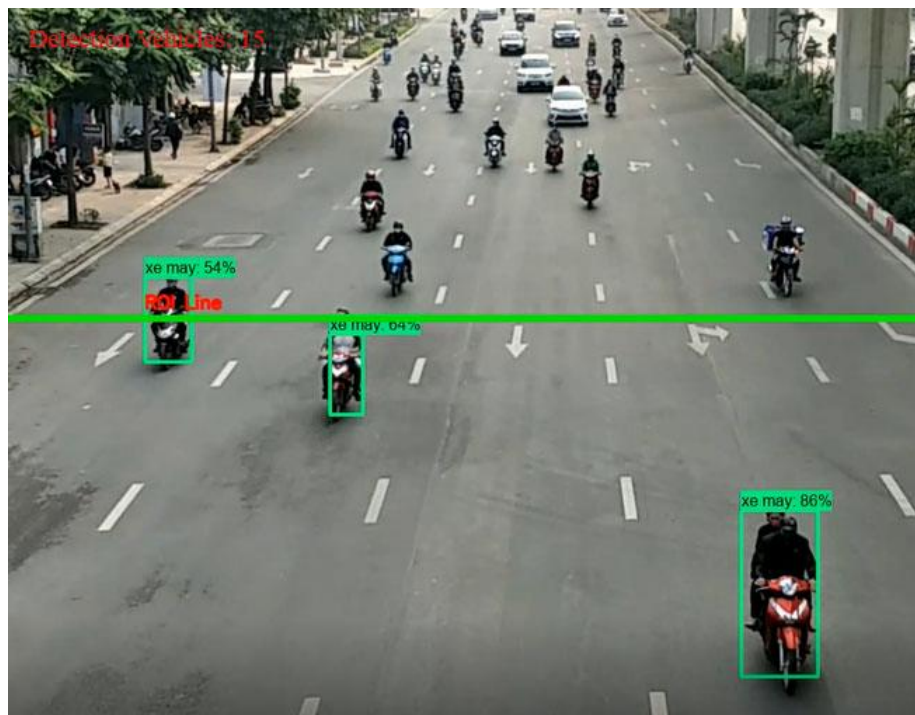


Hình 2.18. Phát hiện và phân loại phương tiện giao thông sử dụng mạng SSD – MobileNet.

2.4.3. Đếm mật độ phương tiện giao thông trong một khu vực giao thông.

Sau khi thực hiện phát hiện và phân lớp phương tiện giao thông tiến hành đếm mật độ phương tiện giao thông. Để đếm mật độ giao thông lưu thông trên

đường tiến hành theo tuần tự các bước sau: Bước đầu tiên, tạo một đường thẳng kẻ ngang (Roi line) qua hai điểm trên làn đường nơi mà cần phải đếm phương tiện giao thông, đường kẻ này phải đáp ứng yêu cầu nằm ở nơi có lượng phương tiện giao thông qua lại nhiều nhất trên đường. Bước tiếp theo là bước đếm số lượng phương tiện giao thông tiến hành như sau khi một phương tiện đi qua đường kẻ ngang thì số lượng phương tiện đếm sẽ tăng lên một đơn vị.



Hình 2.19. Đếm mật độ phương tiện một điểm giao thông trên đường.

2.5. Phát hiện điểm ùn tắc giao thông.

Phát hiện được điểm, nút giao thông có nguy cơ ùn tắc là một công việc có ý nghĩa vô cùng quan trọng. Trong luận văn sử dụng phương pháp đếm mật độ giao thông tại các nút, điểm giao thông theo phương pháp sau đây:

+ Nếu số lượng phương tiện được đếm tại điểm hoặc nút giao thông có giá trị > 50 thì sẽ đưa ra dự đoán tại điểm giao thông đó có nguy cơ xảy ra ùn tắc áp dụng đối với 2 làn đường nhỏ, do nhận thấy với ngưỡng phương tiện này thích hợp bởi vì tại điểm giao thông này có mật độ phương tiện đông đúc.

+ Nếu số lượng phương tiện tại điểm, nút giao thông < 50 , thì đưa ra dự đoán không xảy ra hiện tượng ùn tắc đường.



Hình 2.20. Phát hiện điểm ùn tắc giao thông.

2.6. Kết luận chương 2

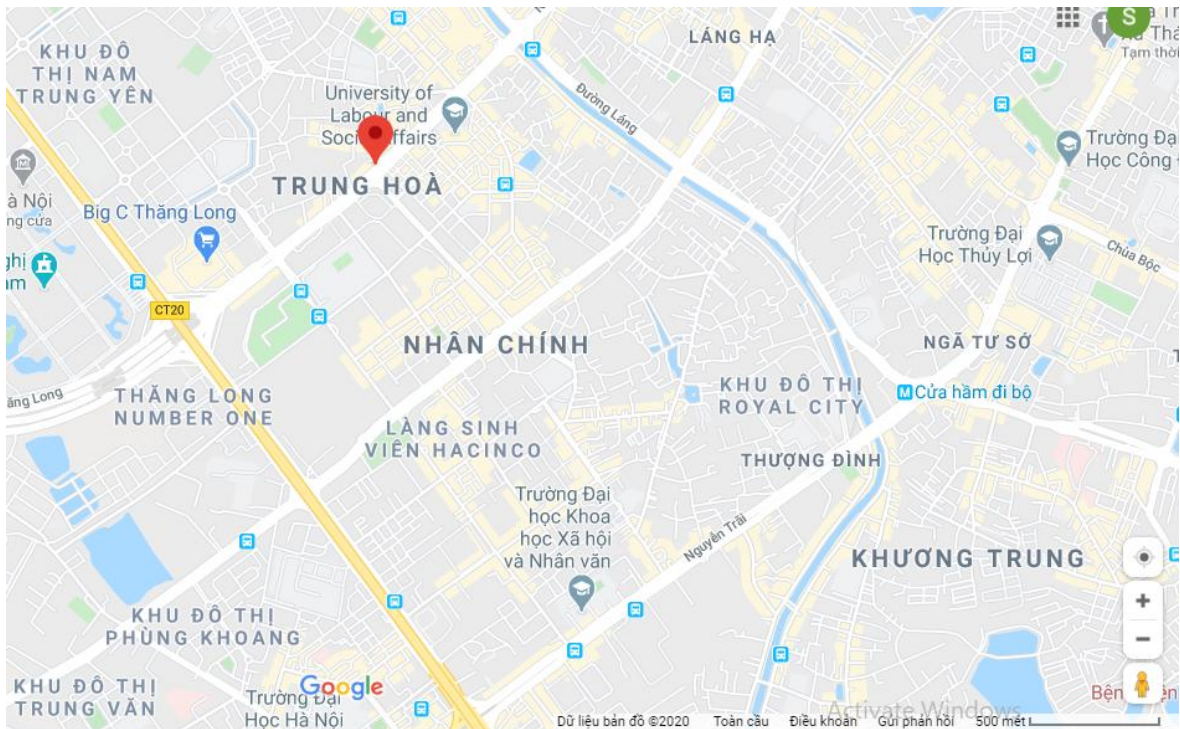
Trong chương này, luận văn đã trình bày về phương pháp phát hiện và phân lớp phương tiện giao thông bằng cách sử dụng mạng SSD – MobileNet trên cơ sở nền mạng nơron tích chập, nhờ cách tính tích chập sâu phân tách nên đã giảm số lượng công việc tính toán đi đáng kể giúp cho tăng tốc độ xử lý của thuật toán. Ngoài ra, cũng đã đề xuất phương pháp đếm mật độ phương tiện giao thông từ đó dự đoán ùn tắc giao thông tại các điểm, nút giao thông trên tuyến đường.

Ở chương kế tiếp, chúng tôi sẽ áp dụng phương pháp này vào thực tế trên tập dữ liệu thu thập được từ phương tiện không người lái UAV trên các tuyến đường Nguyễn Trãi, Khuất Duy Tiến, Trần Duy Hưng và Lê Văn Lương tại thành phố Hà Nội.

Chương 3. THỬ NGHIỆM PHÁT HIỆN ĐIỂM ÛN TẮC GIAO THÔNG DỰA TRÊN TẬP DỮ LIỆU

3.1. Dữ liệu

Tập dữ liệu video thu thập được bằng quay bằng phương tiện không người lái UAV ở thành phố Hà Nội trên các tuyến đường Trần Duy Hưng, Lê Văn Lương, Khuất Duy Tiến và Nguyễn Trãi.



Hình 3.1. Sơ đồ tuyến đường thu thập dữ liệu.

- Dữ liệu được thu thập vào ngày 28/03/2020
- Vị trí không gian: các tuyến đường khảo sát Lê Văn Lương, Nguyễn Trãi, Khuất Duy Tiến và Trần Duy Hưng.
- Thời gian: ban ngày, ban đêm.
- Độ cao: 10m trở lên
- Dạng video: video (.mp4)

Ở hình 3.2 và hình 3.3 là cảnh quay từ UAV bay ở độ cao 25m trong điều kiện thời tiết bình thường trên tuyến đường Trần Duy Hưng và Lê Văn Lương vào buổi sáng lúc 7h00' trở đi. Tuyến đường Trần Duy Hưng có 4 phần đường chia thành 12 làn trên làn đường giữa theo hướng vào trung tâm thành phố với mật độ giao thông rất đông đúc. Còn ở tuyến đường Lê Văn Lương mật độ giao thông thưa thớt hơn, các phương tiện dễ dàng di chuyển trên phần đường của mình.



Hình 3.2. Dữ liệu giao thông thu thập trên tuyến đường Trần Duy Hưng.



Hình 3.3. Dữ liệu giao thông thu thập trên tuyến đường Lê Văn Lương.

Hình 3.4 và Hình 3.5 là 2 video thu thập được trên tuyến đường Khuất Duy Tiến và Nguyễn Trãi. Tuyến đường Khuất Duy Tiến là một trong những tuyến đường thường xuyên ùn tắc giao thông do mật độ giao thông dày đặc và số lượng làn đường ít. Còn tuyến đường Nguyễn Trãi với làn đường rộng phân chia 5 làn nên mật độ giao thông có phần ít hơn.



Hình 3.4. Dữ liệu giao thông thu thập trên tuyến đường Khuất Duy Tiến.



Hình 3.4. Dữ liệu giao thông thu thập trên tuyến đường Nguyễn Trãi.

3.2. Thử nghiệm

3.2.1. Cài đặt môi trường thử nghiệm

Ứng dụng xác định điểm ùn tắc giao thông được tiến hành cài đặt và chạy thử nghiệm trên môi trường Window 10 với những yêu cầu cấu hình sau:

- Khuyến nghị: hệ điều hành từ Window 7 trở lên
- Chip: Intel core i3 3120U trở lên
- Ổ cứng trống 6Gb trở lên
- Card đồ họa (GPU) 2Gb trở lên

Đồng thời cài đặt kèm theo các thư viện sau:

Bảng 3.1: Các thư viện cần cài đặt cho ứng dụng

STT	Tên thư viện
1	Tensorflow object counting API
2	Protobuf 3.0
3	Python - tk
4	Pillow 1.0
5	Lxml
6	Tf Slim
7	Jupyter notebook
8	Matplotlib
9	Tensorflow
10	Cython
11	Contexlib2
12	cocoapi

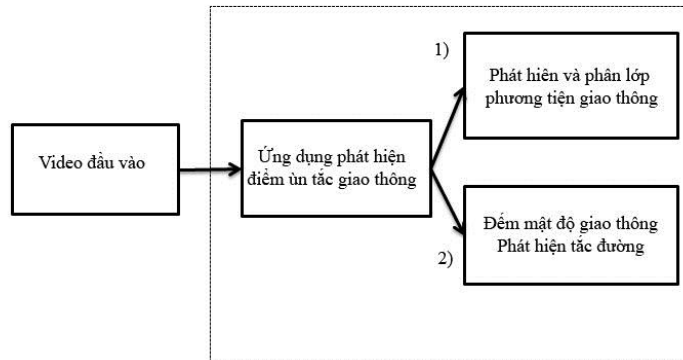
3.2.2. Mô tả hoạt động của ứng dụng

Ứng dụng phát hiện điểm ùn tắc giao thông thực hiện các chức năng sau đây:

- 1) Chức năng phát hiện phương tiện giao thông.

2) Chức năng phân loại và đếm số lượng phương tiện từ đó đưa ra dự đoán tắc đường.

Mô tả các hoạt động của ứng dụng: theo sơ đồ hình 3.5 video đầu vào sẽ được phân tách thành các khung hình. Sau đó tiến hành tiền xử lý ảnh để làm giảm kích thước của ảnh. Tiếp theo sẽ thực hiện phát hiện và phân lớp phương tiện trong ảnh bằng mô hình MobileNet với tập dữ liệu đã được huấn luyện. Sau khi thực hiện xong bước phân lớp phương tiện xong thì tiếp đến sẽ thực hiện đếm mật độ phương tiện giao thông trên đường và đưa ra dự đoán tại điểm, nút giao thông này có ùn tắc hay không.



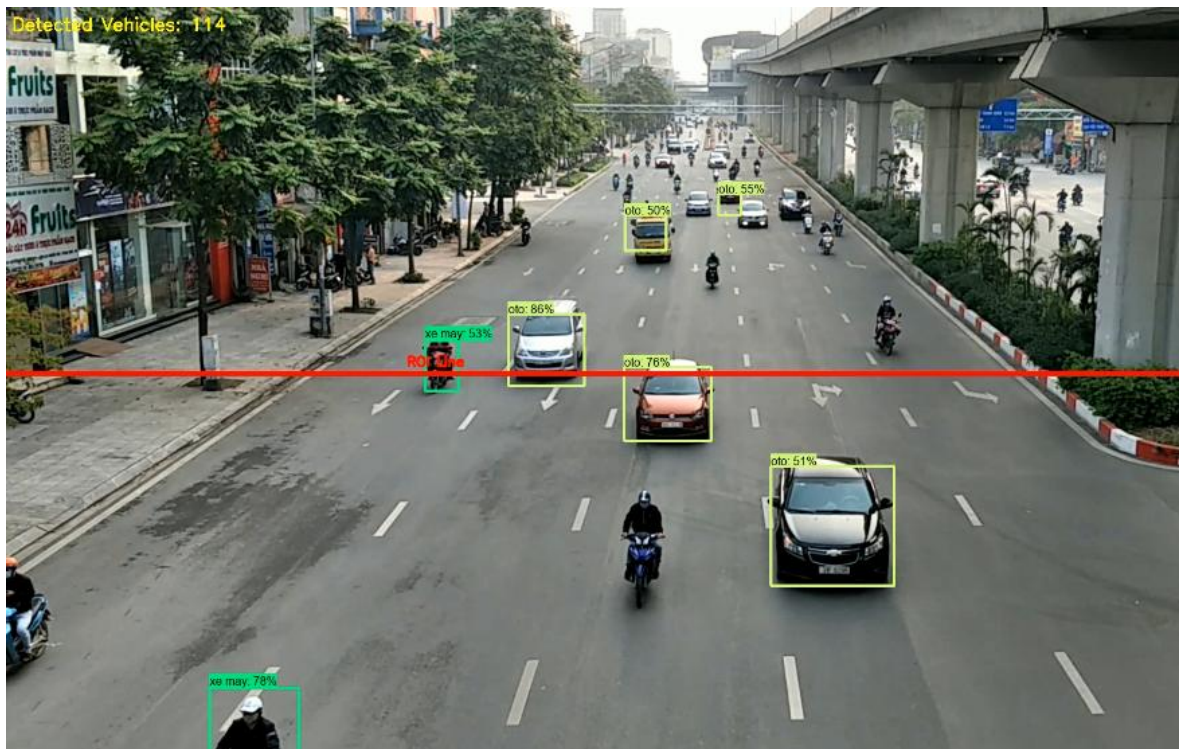
Hình 3.5. Kiến trúc tổng quát ứng dụng phát hiện điểm ùn tắc giao thông.

3.2.3. Kết quả thử nghiệm trên tập dữ liệu thu thập

Kết quả thu được khi thử nghiệm ứng dụng phát hiện ùn tắc giao thông trên video của tuyến đường Nguyễn Trãi thu được kết quả không phát hiện tắc đường tại điểm giao thông này:

Bảng 3.2. Bảng kết quả thử nghiệm trên tuyến đường Nguyễn Trãi.

	Tổng phương tiện	Ô tô	Xe máy	Phân lớp
Thực tế	120	50	70	
Kết quả phát hiện	114	46	68	2
Phần trăm	95%	92%	97%	

**Hình 3.6. Kết quả chạy thử nghiệm video trên tuyến đường Nguyễn Trãi.**

Trên tuyến đường này với số lượng phương tiện đếm được trong ứng dụng là 114 nhưng không dự đoán là ùn tắc bởi vì tuyến đường này có 5 làn xe và mật độ giao thông thưa thớt. Và tỉ lệ chính xác khi thực hiện thuật toán này lên tới 92%.

Bảng 3.3. Bảng kết quả thử nghiệm trên tuyến đường Khuất Duy Tiến.

	Tổng phương tiện	Ôtô	Xe máy	Phân lớp
Thực tế	124	55	69	
Kết quả phát hiện	118	50	68	2
Phần trăm	95.1%	90.9%	98.5%	

**Hình 3.7. Kết quả chạy thử nghiệm trên tuyến đường Khuất Duy Tiến.**

Ở hình 3.7 là một khung hình được trích xuất từ video sau khi xử lý qua ứng dụng ở tuyến đường Khuất Duy Tiến. Thực tế tại điểm giao thông trên hình có tổng 124 phương tiện tiến hành phân ra 2 lớp với độ chính xác 91% đối với xe máy và 98% đối với ô tô. Do tuyến đường này có 3 làn đường và lúc này mật độ lưu thông trên đường tương đối đông nên khi số lượng vượt quá 50 phương tiện thì hệ thống dự báo tắc đường.

3.3. Kết luận chương 3

Ở chương này, luận văn đã tiến hành thực hiện chạy thử nghiệm trên tập dữ liệu thu thập trên 4 tuyến đường có mật độ giao thông đông đúc của Thành phố Hà Nội. Qua đó, thu được kết quả với độ chính xác như mong muốn và thực hiện thời

gian ngắn hơn so với một số mô hình khác. Tuy nhiên, trong quá trình phát hiện điểm ùn tắc giao thông còn có một số hạn chế đó là:

Thứ nhất, đối với tầm bay của UAV từ 5 đến 10 mét việc thu thập các video để xử lý cho kết quả nhận diện và phát hiện các phương tiện cho độ chính xác tương đối cao. Tuy nhiên, đối với tầm bay của UAV trên 25m trở lên thì việc xác định các phương tiện giao thông trở nên khó khăn và không đảm bảo tính chính xác.

Thứ hai, chủ yếu các video được thu thập trong điều kiện thời tiết bình thường và quay vào ban ngày. Nhưng với điều kiện thời tiết xấu, cảnh quay vào ban đêm thì độ chính xác của việc thực hiện mô hình SSD - MobileNet có thể chưa đạt được như mong muốn.

Thứ ba, chưa xác định được vận tốc của các phương tiện giao thông. Trong tương lai, cố gắng thực hiện sử dụng phương pháp đếm cùng với xác định vận tốc phương tiện sẽ nâng cao dự đoán về phát hiện điểm ùn tắc giao thông.

KẾT LUẬN

Luận văn đã khảo sát bài toán phát hiện, phân lớp phương tiện giao thông từ trước đến nay, từ đó đưa ra bài toán riêng cho luận văn là phát hiện điểm ùn tắc giao thông. Dữ liệu trong luận văn thực hiện trên thực tế được thu thập bằng thiết bị bay không người lái. Từ các dữ liệu có được này, cá nhân đã tìm hiểu và nghiên cứu phương pháp hiện điểm ùn tắc giao thông.

Trong luận văn, phương pháp phát hiện điểm ùn tắc giao thông đã trải qua các giai đoạn từ tiền xử lý ảnh. Tiếp đến phát hiện, phân lớp và đếm mật độ phương tiện giao thông sử dụng mô hình SSD - MobileNet để từ đó phát hiện điểm ùn tắc giao thông trên tuyến đường khảo sát trong thành phố Hà Nội với kết quả đạt độ chính xác trên 90%.

Những công việc trên đòi hỏi phải có các nghiên cứu chuyên sâu về các lĩnh vực xử lý hình ảnh, phân tích, thiết kế và xây dựng hệ thống. Tuy nhiên do những hạn chế nhất định về mặt thời gian nên luận văn chỉ mới thực hiện được trên một số tuyến đường trong thành phố Hà Nội. Trong tương lai, sẽ mở rộng thêm các tuyến đường khác trong thành phố Hà Nội, xác định tốc độ của các phương tiện giao thông, thực hiện quay vào ban đêm bằng IR camera và tiến hành cải thiện mô hình SSD - MobileNet để đạt thêm độ chính xác.

DANH MỤC CÁC TÀI LIỆU THAM KHẢO

Tài liệu tiếng việt

- [1] Nguyễn Văn Căn (2015), *Nghiên cứu phát triển một số thuật toán phát hiện và phân loại phương tiện từ dữ liệu video giao thông*, pp 8–10.
- [2] Nguyễn Đắc Thành (2017), *Luận văn nhận dạng và phân loại hoa quả trong ảnh màu*, pp 19-22.

Tài liệu tiếng anh

- [3] Adlin S. T, Kumudha R. (2013), “A Survey on Color Image Enhancement Techniques”, *IOSR Journal of Enginneering (IOSRJEN)*, Vol. 3, Issue 2, pp. 20-22.
- [4] Adu-Gyamfi, Y., Asare, S., Sharma and Titus (2017), “Automated vehicle recognition with deep convolutinal neural networks”, *Transfortation research record journey of the transportation research board*.
- [5] Andrew G.Howard, Menglong Zhu (2011), “Efficient Convolutional Neural Networks for Mobile Vision Applications”.
- [6] Assmaa, O., Mokhtar, K. and Abdelaziz (2013), “Road traffic density estimation using microscopic and macroscopic parameters”, *Image and vision Computing*, Volume 13, No. 11, pp. 887-894.
- [7] Dumoulin, Visin (2016), A guide to convolution arithmetic for deep learning.
- [8] Goncalves, W.N., Machado, B.B and Bruno (2012), “A new method for dynamic texture recognition”, *arXiv preprint arXiv*.
- [9] Huew Enginneering (2015). Introduce to Convolution Neurals Networks – Huew enginneering.
- [10] Joseph Redmon and Ali Farhadi (2016), “Yolo9000: better, faster, stronger”, arXiv: 1612.08242.
- [11] Lempitsky, V. and Zisserman (2010), “Learning to count object in images”, *Advances in neural information processing systems*.
- [12] Mark Sandler, Andrew Howard, Menglong Zhu (2016), “MobileNetV2: Inverted Residuals and Linear Bottlenecks”, arXiv: 1801.04381.
- [13] Ozkurt, C. and Camci (2009), “Automatic traffic density estimation and vehicle

- classification for traffic surveillance systems using neural networks”, *Mathematical and Computational Applications*, Volume 14, No.3, pp 187 – 196.
- [14] Sagar Deb (2005), “Video data management and information retrieval”, University Southern Queensland, pp 20-24.
- [15] Samer, Rishi, Rowen (2015), “Image Recognition Using Convolutional Neural Networks”, Cadence Whitepaper, pp 1-12.
- [16] Uijlings, J.R.R, K.E Van de Sande (2013), “Selective Search for Object Recognition”, *International Journal of Computer Vision*, Vol. 104, No. 2, pp. 154-171.
- [17] Ren, S., He, K., Girshick, R. and Sun (2017), “Towards real-time object detection with region proposal networks”, *IEEE transactions on pattern analysis and machine intelligence*, Volume 39, No. 6, pp. 1137-1149.

Các trang Web

- [18] <https://anninhthudo.vn> , truy cập ngày 20/03/2020
- [19] <https://ladweb.nascom.nasa.gov/> truy cập ngày 12/05/2020
- [20] <https://doisongphapluat.com> truy cập ngày 07/05/2020
- [21] <https://wikipedia.org/wiki/video> truy cập ngày 05/05/2020