

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

---



**NGUYỄN TẤN HẢI**

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

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

**TÓM TẮT LUẬN VĂN THẠC SĨ**

**HÀ NỘI - 2020**

Luận văn được hoàn thành tại:

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

Người hướng dẫn khoa học: TS. NGUYỄN VĂN THỦY

Phản biện 1: .....

Phản biện 2: .....

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại  
Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: ..... giờ ..... ngày ..... tháng ..... .. năm .....

Có thể tìm hiểu luận văn tại:

Thư viện Trường Học viện Công nghệ Bưu chính Viễn thông

## MỞ ĐẦU

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 diễn biến cực kì phức tạp, 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.

# Chương 1. TỔNG QUAN VỀ PHÁT HIỆN ĐIỂM ÛN 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. Các quốc gia trên toàn thế giới nói chung, theo các số liệu thống kê hằng năm tổng thiệt hại mỗi năm do ùn tắc giao thông gây ra cho nền kinh tế của các quốc gia này là cực kỳ to lớn. Còn ở Việt Nam, tại các thành phố lớn ùn tắc giao thông làm cho đất nước của chúng ta thiệt hại hàng tỷ USD mỗi năm theo báo cáo thống kê được 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 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, 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 như sau 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 (oto, xe gắn máy và các phương tiện thô sơ khác) và dự báo tại thời điểm 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 kì 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 faster (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 ùn tắc 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.

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.

Mạng nơron xác định phương tiện giao thông đối với thời tiết bình thường tức là trời nắng đẹp 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 với độ chính xác trên 90%.

### ***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 nơron 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.

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.

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).

### **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.

## **Chương 2. PHƯƠNG PHÁP PHÁT HIỆN ĐIỂM ỨNG 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 ứ tắc giao thông trên tuyến đường khảo sát.

### **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);



***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).

***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 có giá trị lớn hơn 0.

$$f(x) = \max(0, x) \quad (2.1)$$

***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.

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

doán. 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 cồng 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 Feature Map (dựa vào mạng cơ sở VGG16) để tăng hiệu quả trong việc phát hiện thì thường được sử dụng kết hợp sử dụng với ResNet, InceptionNet hoặc MobileNet.
- Áp dụng các bộ lọc tích chập để có thể phát hiện các đối tượng.

## 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.

### 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 (Depthwise 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.

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. Và quan trọng nhất trong mô hình kiến trúc MobileNet 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. 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. Còn Tích chập điểm (Pointwise convolution) sử dụng tích chập 1x1 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:

$$D_k \cdot D_k \cdot M \cdot D_f \cdot D_f \quad (2.2)$$

Công thức tính tích chập điểm là:

$$M \cdot N \cdot D_f \cdot D_f \quad (2.3)$$

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

$$D_k \cdot D_k \cdot M \cdot D_f \cdot D_f + M \cdot N \cdot D_f \cdot D_f \quad (2.4)$$

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

$$D_k \cdot D_k \cdot M \cdot N \cdot D_f \cdot D_f \quad (2.5)$$

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

$$\frac{D_k \cdot D_k \cdot M \cdot D_f \cdot D_f + M \cdot N \cdot D_f \cdot D_f}{D_k \cdot D_k \cdot M \cdot N \cdot D_f \cdot D_f} = \frac{1}{N} + \frac{1}{D_k^2} \quad (2.6)$$

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.

## 2.4. Phương pháp đếm phương tiện giao thông

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].

### 2.4.1. Dữ liệu video kỹ thuật số

a) Định nghĩa 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].

#### b) Đặc trưng của video số

Đặc trưng của video bao gồm: màu, kết cấu, hình dạng và chuyển động.

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 giao thông***

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:

#### ***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.

### 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 dãn” 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.

$$p_x(i) = n_i \quad 0 \leq i < L \quad (2.7)$$

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

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

$$cdf_x = \sum_{j=0}^i p_x(j) \quad (2.8)$$

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

$$y(i) = \text{round}\left(\frac{cdf(i) - cdf_{\min}}{M * N - 1} * (L - 1)\right) \quad (2.9)$$

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

### 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 nơron

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.

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.



### ***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, chúng ta 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.

## **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 tắc đường áp dụng đối với 2 làn đường, do nhận thấy với ngưỡng phương tiện như này mật độ phương tiện đông và dày đặc.

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

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

### **3.1. Tập 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.

- 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)

### **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

#### ***3.2.2. Mô tả cách 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.

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.

### ***3.2.3. Kết quả thử nghiệm***

- 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 phát hiện và phân lớp phương tiện giao thông với độ chính xác trên 90% và không phát hiện tắc đường tại điểm giao thông này.

- Còn ở tuyến đường Khuất Duy Tiến. Thực tế tại điểm giao thông thử nghiệm có tổng 124 phương tiện tham gia giao thông 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.

## 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.