NHẬN DIỆN VÀ PHÂN LOẠI   
BIỂN BÁO GIAO THÔNG

Nguyễn Anh Tuấn   
dept. Computer Vision   
University of Science, VNU   
Ho Chi Minh City, Viet Nam  
19127614@student.hcmus.edu.vn

Nguyễn Hoàng Vũ  
dept. Computer Vision   
University of Science, VNU   
Ho Chi Minh City, Viet Nam  
19127632@student.hcmus.edu.vn

*Abstract*— Phát hiện và nhận dạng biển báo giao thông nắm một vai trò quan trọng trong các hệ thống giao thông thông minh hiện nay, Đồng thời còn đóng góp vào các ứng dụng thực tế như hệ thống xe tự hành, hệ thống hỗ trợ giao thông,.. Báo cáo này sẽ xây dựng một hệ thống nhận diện biển báo giao thông bằng cách xây dựng một mạng CNN đơn giản dựa trên model mạng học sâu VGG.

Keywords: Traffic signs, detection, recognition, Advance Driver Assistance System, CNN

# **GIỚI THIỆU**

Với sự phát triển không ngừng nghỉ của xã hội cũng như sự tiến bộ của khoa học, nhu cầu di chuyển của con người nay ngày càng được lưu tâm. Từ việc đi bộ, cưỡi động vật,... cho đến di chuyển bằng các xe cơ giới. Việc di chuyển bằng xe cơ giới này tạo nên một hệ thống giao thông vô cùng phức tạp.Do đó, “biển báo giao thông” là một trong những thứ đóng vai trò rất quan trọng trong việc truyền tải các nội dung quan trọng một cách nhanh nhất cho các người sử dụng phương tiện (ví dụ: biển báo cấm vượt, quẹo, biển báo giới hạn tốc độ, các biển báo thông báo nguy hiểm,...). Việc này đồng thời mạng lại sự an toàn, đồng thời đảm bào các hệ thống giao thông, cầu đường hoạt động một cách trơn tru khi mà các người lái đều có thông tin về đoạn đường mình di chuyển.

Tuy nhiên, thời đại 4.0 sinh ra một hệ thống cao cấp hơn của giao thông: các hệ thống giao thông thông minh (intelligent transportation system) bao gồm những hệ thống nhúng, giám sát, hỗ trợ giao thông và đồng thời trong tương lai cũng có thể suất hiện một hệ thống nữa đó là: hệ thống xe tự hành. Hai hệ thống này đề cao tính “tự động” trong việc quản lý, di chuyển phương tiện giao thông, do đó, việc khiến cho các hệ thống trên nhận biết được biển báo giao thông là một điều vô cùng cần thiết.

Nhờ vậy, bài toán nhận diện và phân loại biển báo giao thông được ra đời và đang là đề tài nóng hổi cho các nhà nghiên cứu thị giác máy tính và trí tuệ nhân tạo tìm tòi, nghiên cứu. Bài báo cáo này sẽ nêu lên phương pháp xây dựng một hệ thống nhận diện biển báo thông qua một mạng neuron tích chập CNN sử dụng model VGG, một vài ví dụ thực tế cũng như một vài cải tiến nhỏ để phát triển hệ thống.

# **TỔNG QUAN**

## VẤN ĐỀ VÀ BÀI TOÁN (PROBLEM STATEMENT)

Bài toán phân loại biển báo giao thông cũng giống như các bài toán phân loại khác, ẩn trong nó là 2 bài toán nhỏ:

Thứ nhất: bài toán dò tìm đối tượng (Object Detection), ở bước này, ta sẽ “tìm” vật thể, ở đây là các biển báo giao thông. “Tìm” ở đây chính là xác định vị trí tọa độ trên ảnh đồng thời xác định kích thước của vật thể đó.

Thứ hai: bài toán nhận diện (Object Recognition), bài toán này hướng đến việc sẽ đi phân lớp các vật thể vào các lớp của nó. Ở đây, chúng ta sẽ phân loại xem các biển báo mà ta tìm được ở bài toán 1 là biển báo loại gì.

## CÁC NGHIÊN CỨU LIÊN QUAN (RELATED WORK)

Từ trước tới nay, đã có rất nhiều nghiên cứu về việc nhận diện được biển báo giao thông:

Việc phát hiện biển báo giao thông thường dựa trên các đặc trưng hình dạng và màu sắc.

Các biển báo giao thông đa phần đều tuân theo một bảng màu đã được quy định. Do đó, việc sử dụng màu như một đặc trưng để dò tìm các biển báo là một trong những hướng đi lâu đời của bài toán. Các phương pháp phát hiện biển báo giao thông dựa trên màu sắc (Color) là các phương pháp lâu đời và dễ hiểu **[2].** Tuy nhiên thì do hệ màu RGB khá nhạy cảm với các yếu tố khác như: ánh sáng, nhiễu, blur,... Đã có nhiều nghiên cứu đã được đề ra như: sử dụng hệ màu HSV**[6]**, YUV, HSI,… Dù đã cải tiến được đôi chút, song, các đặc trưng màu này vẫn quá bị ảnh hưởng bởi các yếu tố khác như: thời tiết, ánh sáng, …

Song song cùng với đặc trưng màu, đặc trưng dáng (Shape) cũng được dùng để nhận diện biển báo. Các biển báo thường có hình dáng cố định đặc trưng như: Tam giác, vuông, lục giác, tròn,… Dựa trên việc này, nhiều nghiên cứu đã loại bỏ hoàn toàn đặc trưng màu và thay vào đó là các đặc trưng dáng để nhận dạng **[7].**

Ở những thời kỳ đầu của bài toán, việc phân loại các đặc trưng này thường được dựa trên các thuật toán phân lớp như: Random Forest **[1]**, SVM **[2]**, KNN, …Các phương pháp trên đều có thể hoạt động, nhưng vẫn có nhiều hạn chế.

Bài toán này trở nên đột phá khi mà được kết hợp với mạng neuron tích chập CNN. Từ khi áp dụng CNN vào, đã có vô số các nghiên cứu để đưa ra các model, các thuật toán phù hợp với việc phân loại biển báo giao thông. **[4]** sử dụng model VGG đồng thời cải tiến nó thành một model mới có tên IVGG (improved VGG), tuy nhiên phương pháp này lại có nhược điểm là có tỉ lện nhận dạng thấp với các ảnh có nền tối, mờ. **[5]** Tự xây dựng một mô hình CNN đơn giản, đồng thời tinh chỉnh các mô hình như Faster RCNN và YOLOv4, báo cáo cũng cho thấy được sự so sánh giữa hiệu năng các phiên bản YOLO từ v1 – v4 với nhau. **[8]** xây dựng một quy trình dựa trên phân loại đặc trưng màu , sừ dụng HOG kết hợp với mạng CNN, để phát hiện và nhận dạng các biển báo giao thông, giúp đạt được độ chính xác phân loại và tốc độ tính toán tốt hơn. Tuy nhiên dù mạng lại kết quả rất khả quan trong độ chính xác khi nhận dạng biển báo, việc sử dụng các thuật toán deep learning khiến cho thời gian tính toán lâu hơn, dẫn đến một vấn đề khá lớn trong việc áp dụng vào real-time, thực tế. Vấn đề này vẫn luôn được nghiên cứu cũng như đưa ra các giải pháp nhằm cải thiện xuyên suốt thời gian gần đây.

# **PHƯƠNG PHÁP**

## Sơ đồ tổng quát

## Tiền xử lý và phân chia dữ liệu

*Phương pháp Phân chia dữ liệu:*

* Đây là bước đầu tiên việc phân chia: load dữ liệu và xử lý dữ liệu training. Chia tập dữ liệu ra thành các tập như: Training Set – dùng để huấn luyện máy học.
* Testing Set - là tập dữ liệu dùng để test sau khi máy đã học xong
* Validation Set - test cũng giống như tập training set, nó cũng bao gồm các cặp giá trị input và ouput tương ứng. Nhưng nó lại khác training set ở chỗ, nó được sử dụng để kiểm thử độ chính xác của mô hình máy học trong quá trình huấn luyện

Lưu ý: Testing Set được dùng để kiểm thử sau quá trình huấn luyện, còn Validation Set được sử dụng để kiểm thử trong quá trình huấn luyện.

Tiếp đó sẽ là bước xử lý dữ liêu, cơ bản làm bước dùng để biến đổi ảnh đầu vào thành một ảnh phù hợp hơn, bằng các cách như resize ảnh, blur ảnh, ...

Thương các dữ liệu sẽ được chia theo tỉ lệ: 50% training – 25% testing – 25% validation.

*Tiền xử lý ảnh:*

Để giúp mô hình phân loại dễ dàng hơn, trước khi phân loại, các ảnh cần được xử lý trước. Có rất nhiều phương pháp tiền xử lý có thể được áp dụng như: greyscale, thresholding, histogram equalization,... Do việc các biển báo giao thông được đặt ngoài trời, bị ảnh hưởng nhiều bởi các ngoại cảnh như: ánh sáng, mờ,... Nên việc sử dụng histogram equalization giúp cân bằng lại các kênh màu sẽ khiến cho việc xử lý ảnh tốt hơn.

## Object Detection sử dụng Contour

Contour được hiểu đơn giản là một đường cong liên kết toàn bộ các điểm liên tục (dọc theo đường biên) mà có cùng màu sắc hoặc giá trị cường độ. Contour rất hữu ích trong phân tích hình dạng, phát hiện vật thể và nhận diện vật thể. Để có thể tìm được các contour của ảnh, ta phải tìm được các biên cạnh, qua các thực nghiệm, nhóm đã chọn thuật toán Canny để sử dụng cho việc tìm biên cạnh này vì sự chính xác cao và tốc độ sử lý nhanh của thuật toán, đồng thời các cạnh của Canny cũng rất mỏng do đó dễ tìm được Contour hơn.

Sau khi tìm hiểu, nhóm nhận thấy các biển báo giao thông đa phần là hình vuông, tròn, chữ nhật, tam giác,... Do đó, nhóm nảy ra ý tưởng, thay vì sử dụng một model để tìm kiếm và đánh dấu, nhóm sẻ chỉ sử dụng contour cho phân đoạn này nhằm tiết kiệm thời gian sử lý.

Sau khi dùng Contuor để tìm được các không gian ảnh được cho là biển báo, nhóm sẽ dùng bounding box để crop thành các ảnh nhỏ hơn, sau đó đưa các ảnh này vào mô hình VGG để nhận diện. Việc này mang lại rất nhiều lợi ích cho việc nhận diện vật thể, giúp cho máy loại bỏ các background không cần thiết, đồng thời có thể nhận diện được nhiều vật thể trong một ảnh thay vì chỉ trả ra một giá trị.

## Mô hình VGG

### Sơ lược về VGG

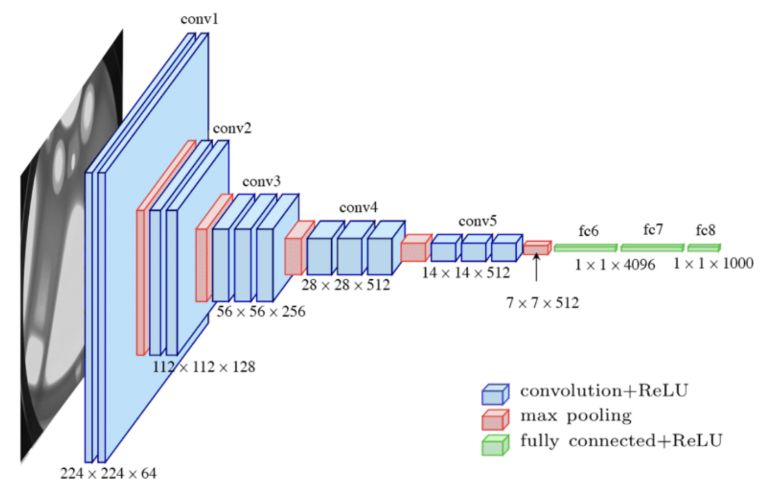
VGG là mô hình đã chiên thắng của cuộc thi ImageNet năm 2014, được phát triển dựa trên mô hình AlexNet. Mô hình VGG này có hai đặc điểm nổi bật:

Thứ nhất: các kernels dùng cho tích chập thường rất nhỏ, đa phần là 3x3, có một số là 1x1. Các phép tích chập thường đi kèm với các hàm kích hoạt (activition function) như Sigmoid, Tanh, ReLU,... nên rất dễ tùy biến thành các hàm phong phú hơn.

Thứ hai: các kernels dùng cho việc pooling cũng khá nhỏ, các kernels này thường có kích thước là 2x2, nhỏ hơn so với kernels 3x3 của AlexNet. Việc này khiến cho các lớp pooling sâu hơn và ma trận lớp đầu vào (feature map) rộng hơn.

Với việc kernels tích chập tập trung vào việc mở rộng số lượng các kênh, đồng thời các kernels pooling thu hẹp chiều ngang và cao của các lớp, sẽ khiến cho cấu trúc của mô hình sâu và rộng hơn.

### Mô hình VGG16



Mô hình VGG16 là một trong những mô hình VGG cổ điển nhất. Với VGG16: ảnh đầu vào sẽ có kích thước (224x224x3), có 13 lớp tích chập gồm: lớp conv1 và conv2 có 2 lớp tích chập; conv3, conv4, conv5 sẽ có 3 lớp tích chập. Sau mỗi lớp tích chập sẽ có 1 lớp pooling giúp giảm kích thước lớp đó xuống một nữa so với kích thước ban đầu. Và ở phần cuối ta có 3 lớp fc6, fc7, fc8 mỗi lớp chứa một lớp fully-connected và lớp cuối là một lớp dùng để phân lớp.

## Đánh giá mô hình

Việc đánh giá mô hình thường được thực hiện trên dữ liệu mà mô hình chưa từng được học – trên validation set và test set và ở báo cáo này cũng vậy.

Độ đo mà báo cáo dùng để đánh giá mô hình:

*Accuracy:*

Accuracy (độ chính xác) là một độ đo hết sức cơ bản. Độ chính xác là tỉ lệ giữa số điểm dữ liệu được dự đoán đúng và tổng số điểm dữ liệu.

Nếu ta có tập các lớp là N (ở báo cáo này là 43), Accuracy sẽ bằng trung bình cộng của tất các các Accuracy của N lớp:

Tuy nhiên, một mô hình có độ chính xác cao chưa hẳn đã tốt. Accuracy lộ rõ hạn chế khi được sử dụng trên bộ dữ liệu không cân bằng (imbalanced dataset), khi các tập dự liệu của từng lớp không đồng đều với nhau, lớp thì quá nhiều, lớp thì quá ít.

# **THỰC NGHIỆM**

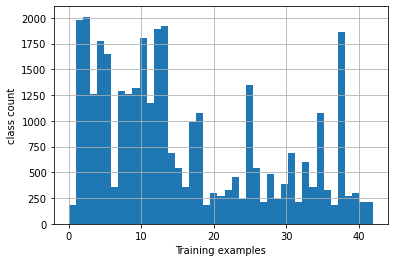
## Tập dữ liệu (Dataset Collection)

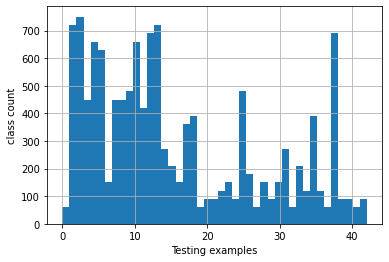
Sử dụng dataset là: German Traffic Sign Recognition Benchmark, [GTSRB](https://www.kaggle.com/meowmeowmeowmeowmeow/gtsrb-german-traffic-sign?select=Train).

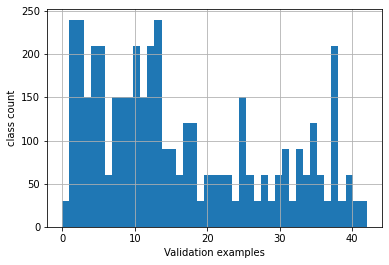
Được sử dụng lần đầu tại IJCNN năm 2011,bộ dữ liệu này bao gồm khoảng 51839 hình ảnh được lưu trữ dưới định dạng png chia làm 43 lớp khác nhau theo chủ đề biển báo giao thông. Trong đó có 39,209 được dùng là ảnh dùng cho (training set) và 12,630 ảnh dùng để test và thẩm định (test set và validation set). Các ảnh có kích thước từ khoảng 15x15 cho đến 250x250 pixels .Các ảnh trong tập dữ liệu đều có điều kiện ánh sáng, khoảng cách, thời tiết phong phú, đa dạng, giúp cho việc tranning và testing tốt hơn. Tập dữ liệu có một thư mục train chứa các hình ảnh bên trong mỗi lớp và một thư mục test chứa các hình ảnh dùng để thử nghiệm mô hình.

Để thuận tiện cho việc lưu trữ, các ảnh trong dataset đã được mã hóa bằng thư việc **pickle** của python, giúp giảm mạnh kích thước của các tập trong việc lưu trữ.

*Biểu đồ phân bố dữ liệu của các tập:*

**

**

**

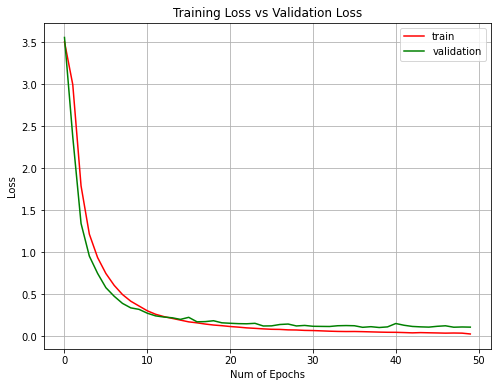
## Môi trường thực nghiệm

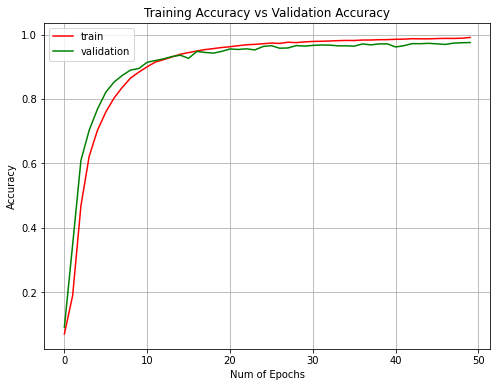
Nhóm sử dụng GoogleColab để train mô hình VGG

## Thực nghiệm – Huấn luyện Model

*Số lượng Epochs*: 50

*Biểu đồ Hàm Loss và Accuracy của các tập training và validation qua từng Epoch*:





*Độ đo Accuracy*:

Accuracy của lớp No.1 : 86.67%

Accuracy của lớp No.2 : 99.31%

Accuracy của lớp No.3 : 99.73%

Accuracy của lớp No.4 : 93.33%

Accuracy của lớp No.5 : 98.03%

Accuracy của lớp No.6 : 92.54%

Accuracy của lớp No.7 : 90.67%

Accuracy của lớp No.8 : 95.78%

Accuracy của lớp No.9 : 98.89%

Accuracy của lớp No.10 : 100.00%

Accuracy của lớp No.11 : 99.09%

Accuracy của lớp No.12 : 98.10%

Accuracy của lớp No.13 : 99.57%

Accuracy của lớp No.14 : 99.58%

Accuracy của lớp No.15 : 100.00%

Accuracy của lớp No.16 : 100.00%

Accuracy của lớp No.17 : 100.00%

Accuracy của lớp No.18 : 98.06%

Accuracy của lớp No.19 : 82.31%

Accuracy của lớp No.20 : 100.00%

Accuracy của lớp No.21 : 95.56%

Accuracy của lớp No.22 : 64.44%

Accuracy của lớp No.23 : 82.50%

Accuracy của lớp No.24 : 93.33%

Accuracy của lớp No.25 : 97.78%

Accuracy của lớp No.26 : 97.08%

Accuracy của lớp No.27 : 89.44%

Accuracy của lớp No.28 : 51.67%

Accuracy của lớp No.29 : 94.67%

Accuracy của lớp No.30 : 92.22%

Accuracy của lớp No.31 : 90.67%

Accuracy của lớp No.32 : 100.00%

Accuracy của lớp No.33 : 98.33%

Accuracy của lớp No.34 : 99.05%

Accuracy của lớp No.35 : 99.17%

Accuracy của lớp No.36 : 98.72%

Accuracy của lớp No.37 : 99.17%

Accuracy của lớp No.38 : 98.33%

Accuracy của lớp No.39 : 97.25%

Accuracy của lớp No.40 : 96.67%

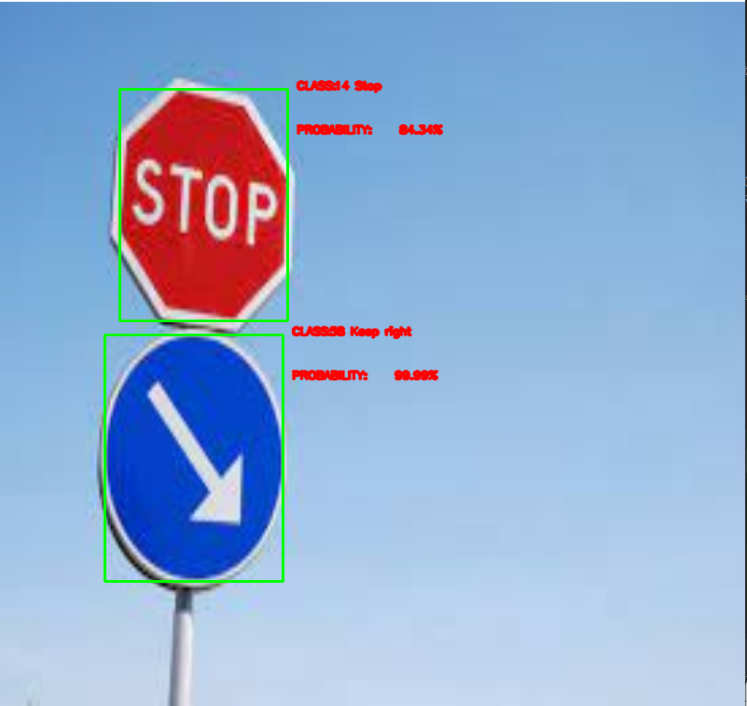
Accuracy của lớp No.41 : 86.67%

Accuracy của lớp No.42 : 68.33%

Accuracy của lớp No.43 : 100.00%

Accuracy: 96.33%

## Kết quả







# **KẾT LUẬN**

Trong bài báo này, nhóm đã sử dụng một kỹ thuật thị giác máy tính là phát hiện Contour cho bước Detection ( Dò tìm), kết hợp cùng mô hình VGG cho bước Classification ( Phân lớp) từ đó giải quyết được bài toán nhận diện biển báo giao thông. Tuy đã thuật toán đã chạy hoàn chỉnh, xong vẫn còn một số khó khăn như:

* Hoạt động không tốt ở real-time khi các biển báo khá nhỏ và bị mờ.
* Việc sử dụng Contour có một nhược điểm lớn là các biển báo thường bị biến dạng khi sử dụng ở Real-time.
* Vẫn bị ảnh hưởng nhiều bởi các yếu tố ngoại cảnh.

Một yếu tố nữa đó chính là tập dữ liệu GTSRB dù là một tập dữ liệu rất tốt nhưng lại có sự phân bố số lượng ảnh của từng lớp không đồng đều, việc này cũng khá ảnh hưởng đến việc huấn luyện mô hình cũng như đánh giá mô hình.

Nhóm cũng đưa ra một vài phương pháp để cải tiến bài toán như:

* Áp dụng mô hình YOLO, một mô hình one-stage thích hợp cho việc real-time, hoạt động rất nhanh, nhưng đánh đổi lại là sẽ giảm độ chính xác.
* Tìm, cải thiện tập dữ liệu sao cho đồng đều và phong phú hơn giữa các tập dữ liệu.
* Tìm một phương pháp tốt hơn cho phần detection như: Segmation, sử dụng model để detect,...

##### **References**

1. Jack Greenhalgh and M. Mirmehdi, [“Traffic sign recognition using MSER and Random Forests”](https://ieeexplore.ieee.org/abstract/document/6333901)
2. David Soendoro and Iping SuprianaTraffic, [Sign Recognition with Color-based Method, Shape-arc Estimation and SVM](https://ieeexplore.ieee.org/document/6021584)
3. Faming Shao, Xinqing Wang, Fanjie Meng, Ting Rui, Dong Wang, and Jian Tang, [Real-Time Traffic Sign Detection and Recognition Method Based on Simplified Gabor Wavelets and CNNs](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6210476/)
4. Shuren Zhou, Wenlong Liang, Junguo Li and Jeong-Uk Kim, [Improved VGG Model for Road Traffic Sign Recognition](https://www.techscience.com/cmc/v57n1/22953)
5. Njayou Youssouf, [Traffic Sign Detection and Recognition with](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4012734)

[FasterRCNN and YOLOV4](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4012734)

1. Hasan Fleyeh, Syed Omer Gilani[, Road sign detection and recognition using fuzzy artmap: A case study swedish speed-limit signs.](https://www.researchgate.net/publication/220909199_Road_sign_detection_and_recognition_using_fuzzy_artmap_A_case_study_swedish_speed-limit_signs)
2. Pavel Paclíck, Jana Novovicova, [Road Sign Classification without Color Information](https://www.semanticscholar.org/paper/Road-Sign-Classification-without-Color-Information-Pacl%C3%ADk-Novovicov%C3%A1/a0d1f3078208fb101e66d54765f86aeb8d606678)
3. Ali Youssef, Dario Albani, Daniele Nardi, and Domenico D. Bloisi, [Fast Traffic Sign Recognition Using Color Segmentation and Deep Convolutional Networks](https://www.semanticscholar.org/paper/Fast-Traffic-Sign-Recognition-Using-Color-and-Deep-Youssef-Albani/d1cc94c082313dc31f8885f003334cb4283c3d04)