PHÁT HIỆN, NHẬN DẠNG BIỂN BÁO GIAO THÔNG DỰA TRÊN THUẬT TOÁN CNN VÀ SVM

Đoàn Lê Anh Duy, Lâm Hà Yến, Cao An Gia Lộc, Đoàn Quang Anh

Khoa Công nghệ Thông tin, Đại học Nông Lâm TP. HCM

Tóm tắt. Biển báo giao thông là những biển báo được dựng ven đường nhằm mục đích cung cấp thông tin đến người tham gia giao thông. Biển báo bao gồm các hình ảnh, ký tự được tiêu chuẩn hóa và đơn giản hóa giúp cho việc lưu thông quốc tế dễ dàng hơn (giảm bớt rào cản ngôn ngữ) cũng như giúp tăng cường an toàn giao thông. Tuy đã được tiêu chuẩn hóa và đơn giản hóa nhưng việc nhận biết biển báo vẫn còn là một vấn đề khó khăn đối với nhiều cá nhân. Do đó, trong bài báo này chúng tôi sẽ sử dụng thuật toán CNN và SVM giúp mọi người giải quyết vấn đề phát hiện và nhận biết nhiều biển báo giao thông không chồng lắp với nhau trong môi trường ánh sáng bình thường. Hai thuật toán CNN và SVM sử dụng dữ liệu đầu vào là 17909 hình ảnh biển báo giao thông hình tròn của nước Đức với 18 loại biển báo khác nhau. Thuật toán CNN sẽ được chúng tôi xây dựng mô hình học sâu với kiến trúc mạng tích chập gồm 9 khối mạng tích chập và khối cuối là đầu ra softmax. Thuật toán SVM sử dụng từ thư viện Sklearn kết hợp với phương pháp rút trích đặc trưng HOG. Kết quả cho thấy thuật toán CNN có tốc độ nhận dạng chậm hơn thuật toán SVM nhưng cho ra kết quả chính xác hơn SVM.

1. GIỚI THIỆU

Song hành cùng với sự phát triển của các thành tựu khoa học kỹ thuật hiện đại và không ngừng nâng cao đời sống văn hóa thì vấn đề giao thông cũng từng bước được cải thiện và phát triển mạnh mẽ. Không chỉ ở Việt Nam mà giao thông ở các nước trên thế giới đang là vấn đề nóng cần được quan tâm, bởi sự gia tăng về tai nạn giao thông đường bộ, nguyên nhân phần lớn là do người tham gia giao thông không làm chủ tốc độ, không chấp hành hiệu lệnh, không quan sát hoặc kịp nhận ra các biển báo và tính hiệu giao thông. Vì thế, việc xây dựng hệ thống nhận diện biển báo giao thông hỗ trợ nhận diện nội dung của từng biển báo, giúp người tham gia giao thông điều khiển phương tiện đi theo đúng quy định, để đảm bảo an toàn, góp phần hạn chế những tai nạn giao thông và giảm thiểu những hậu quả sau tai nạn. Ngoài ra hệ thống nhận diện biển báo giao thông còn là một yếu tố quan trọng trong việc tạo ra các loại xe tự hành là xu hướng thiết yếu của xã hội phát triển.

Hệ thống nhận diện biển báo giao thông sử dụng ba thuật toán là CNN, SVM và HOG.

CNN (Convolutional Neural Network) là một trong những mô hình Deep Learning tiên tiến được dùng trong trong nhiều bài toán như nhận dạng ảnh, phân tích video, ảnh MRI, hoặc cho các bài của lĩnh vự xử lý ngôn ngữ tự nhiên,và hầu hết đều giải quyết tốt các bài toán này, CNN sẽ so sánh hình ảnh dựa theo từng mảnh và những mảnh này được gọi là Feature, CNN bao gồm những phần lớp cơ bản là: convolutional layer, relu layer, pooling layer, fully connected layer, cấu trúc cơ bản của CNN bao gồm 3 phần chính là: local

receptive field (trường cục bộ), shared weights and bias (trọng số chia sẻ), pooling layer (lớp tổng hợp).

SVM (Support Vector Machine) là 1 thuật toán học máy thuộc nhóm Supervised Learning (học có giám sát) cho kết quả cao, ổn định, chịu đựng nhiễu tốt được sử dụng trong các bài toán phân lớp dữ liệu (classification) hay hồi qui (regression).

HOG (histrogram of oriented gradients) là thuật toán tuy cổ điển nhưng cũng rất hiệu quả trong xử lý ảnh, thuật toán này tạo ra các bộ mô tả đặc trưng nhằm mục đích phát hiện vật thể. Từ một bức ảnh, ta sẽ lấy ra 2 ma trận quan trọng giúp lưu thông tin ảnh đó là độ lớn gradient (gradient magnitute) và phương của gradient (gradient orientation). Bằng cách kết hợp 2 thông tin này vào một biểu đồ phân phối histogram, trong đó độ lớn gradient được đếm theo các nhóm bins của phương gradient. Cuối cùng ta sẽ thu được véc tơ đặc trưng HOG đại diện cho histogram. Sơ khai là vậy, trên thực tế thuật toán còn hoạt động phức tạp hơn khi véc tơ HOG sẽ được tính trên từng vùng cụ bộ như mạng CNN và sau đó là phép chuẩn hóa cụ bộ để đồng nhất độ đo. Cuối cùng véc tơ HOG tổng hợp từ các véc tơ trên vùng cục bộ.

Hai thuật toán CNN và SVM sử dụng dữ liệu đầu vào là 17909 hình ảnh biển báo giao thông hình tròn của nước Đức với 18 loại biển báo khác nhau. Kết quả cho thấy thuật toán CNN có tốc độ nhận dạng chậm hơn thuật toán SVM nhưng cho ra kết quả chính xác hơn SVM.

2. CÁC CÔNG TRÌNH LIÊN QUAN

1.1. Phạm Nguyên Khang, Trần Nguyễn Minh Thư, Đỗ Thanh Nghị, 2017. Điểm danh bằng mặt người với đặc trưng GIST và máy học véc-tơ hỗ trợ.

Hệ thống điểm danh bằng mặt người thực hiện rút trích tự động khuôn mặt người trong ảnh thu được từ camera và xác định danh tính của đối tượng trong hệ thống dựa vào nội dung của ảnh khuôn mặt rút trích được. Hệ thống bao gồm 2 bước chính là định vị khuôn mặt dựa trên các đặc trưng Haar-like kết hợp với mô hình phân tầng (Cascade of Boosted Classifiers - CBC) và định danh đối tượng từ ảnh khuôn mặt sử dụng mô hình máy học SVM. Dữ liệu huấn luyện được lấy từ tập dữ liệu 6722 ảnh của 132 đối tượng là những sinh viên khoa CNTT-TT Trường Đại học Cần Thơ. Kết quả mô hình đạt 99.29% độ chính xác trên tập kiểm tra.

Định vị khuôn mặt với vô mình phân tần CBC sử dụng đặc trưng Haar-like. Định vị khuôn mặt bằng cách di chuyển cửa sổ trượt trên ảnh (từ trái sang phải, từ trên xuống dưới), rút trích đặc trưng Haar-like của vùng ứng viên (cửa sổ đang xét), đưa vào mô hình phân lớp Adaboost [Freund & Schapire, 1995] theo thứ tự từ tầng 1 đến tầng thứ N của mô hình phân tầng CBC, nếu ở tầng thứ t mô hình h_t phân lớp vùng ứng viên không phải là khuôn mặt người thì cửa sổ trượt tiếp tục đến vị trí tiếp theo trên ảnh, nếu vùng ứng viên được mô hình ht ở tầng thứ t phân lớp là khuôn mặt người thì vùng ứng viên tiếp tục chuyển đến tầng thứ $t+1$ để xét vùng ứng viên có phải mặt người hay không, quá trình tiếp tục cho đến

khi tầng N mô hình h_N phân lớp là khuôn mặt người thì vùng ứng viên được xác định là khuôn mặt người.

Rút trích đặt trưng GIST từ ảnh khuôn mặt và thực hiện định danh đối tượng bằng máy học véc-tơ hỗ trợ SVM. Trước tiên, ảnh khuôn mặt cần chuẩn hóa về kích thước 128x128 và chuyển sang bước:

Tiền sử lý: tách 3 kênh màu (đỏ, xanh lá cây, xanh dương), biến đổi tỷ lệ lôgarit (giảm chênh lệch tỷ lệ các điểm ảnh tối so với các điểm ảnh sáng), thêm các điểm ảnh biên, lọc trắng ảnh (cân bằng năng lượng phổ), lọc chuẩn hóa độ tương phản, xóa các điểm ảnh biên. Sinh 20 bộ lọc Gabor để áp dụng lên từng ảnh (đỏ, xanh lá cây, xanh dương, đã được tiền xử lý và chuyển về miền tần số).

Chia ảnh thành 16 vùng riêng biệt bằng nhau, tính giá trị trên mỗi vùng bằng cách lấy tổng giá trị của các điểm ảnh trên mỗi vùng chia cho số điểm ảnh của vùng, thực hiện lần lượt 20 bộ lọc trên 16 vùng của ảnh đỏ, xanh lá cây, xanh dương, thu được véc-tơ có 960 thành phần (chiều).

Sau khi rút trích đặc trưng GIST từ ảnh khuôn mặt tạo ra tập dữ liệu có m dòng (ảnh khuôn mặt), mỗi dòng có 960 chiều (đặc trưng GIST) và nhãn (đối tượng trong hệ thống) sẽ được đưa vào máy học véc-tơ hỗ trợ SVM để tiến hành định danh.

Ưu điểm:

- Thời gian huấn luyện nhanh (6722 ảnh / 4.65 phút).
- Độ chính xác cao lên đến 99.29%.

Nhược điểm:

- Thời gian nhận dạng lâu (1.95 phút / người).
- Khó nhận dạng khi ở môi trường thiếu sáng hoặc camera kém chất lượng do không có bước cải thiện chất lượng hình ảnh.

1.2. Trương Quốc Bảo, Trương Hùng Chen, Trương Quốc Định, 2015. Phát hiện và nhận dạng biển báo giao thông đường bộ sử dụng đặc trưng HOG và mạng nơron nhân tạo

Hệ thống phát hiện nhận dạng biển báo giao thông đường bộ sử dụng đặc trưng cục bộ HOG và mạng nơron nhân tạo có khả năng phát hiện và nhận dạng hầu hết các loại biển báo giao thông như biển báo cấm, biển báo nguy hiểm, biển báo hiệu lệnh và biển chỉ dẫn không bị chồng lắp. Thực nghiệm được tiến hành với 31 video với thời gian trung bình để phát hiện và nhận dạng các biển báo giao thông trên một frame ảnh xấp xỉ 0.021 giây khi sử dụng mô hình phân lớp với mạng nơron nhân tạo và khoảng 0.099 giây khi dùng mô hình phân lớp SVM và độ chính xác nhận dạng khoảng 94%.

Phân đoạn ảnh: trong nghiên cứu này phân đoạn ảnh dựa vào màu đỏ (Red) trên các biển báo cấm và nguy hiểm; màu xanh lam (Blue) trên các biển hiệu lệnh và chỉ dẫn. Đầu tiên,

ảnh đầu vào trong không gian màu RGB được chuyển sang không gian màu IHLS. Sau khi chuyển ảnh sang không gian màu IHLS, giá trị H và S được chọn tương ứng với màu đỏ hoặc màu xanh lam trên biển báo giao thông. Đối với màu đỏ, những điểm ảnh (pixels) có giá trị H<15 hoặc H>183 và S>16 được thể hiện trong ảnh trắng đen với màu trắng (giá trị 1), những điểm ảnh còn lại được thể hiện với màu đen (giá trị 0). Đối với màu xanh lam, tương tự như trên, những pixels có giá trị 143<H<170 và S>36 được thể hiện bằng màu trắng, những pixels còn lại được thể hiện bằng màu đen. Sử dụng ràng buộc tỉ lệ khung hình chiều rộng *w* và chiều cao *h* thỏa *w/h* < 1/3 và *h/w* < 1/7 để chọn chính xác vùng ứng viên.

Phát hiện vùng ứng viên: ảnh trắng đen thu được ở giai đoạn phân đoạn ảnh được lọc bằng bộ lọc Median kích thước 5x5 để loại bớt các vùng nhiễu. Tiếp theo, sử dụng hàm *findContours* trong thư viện OpenCV để dò biên của các đối tượng trong ảnh. Do biển báo giao thông là các đa giác lồi nên sử dụng hàm *isContourConvex* để tìm các đa giác lồi đó.

Đặc trưng HOG và thực hiện nhận dạng bằng mạng nơron nhân tạo (ANNs): HOG sẽ dựa vào hình dạng và trạng thái của vật có thể rút trích đặc trưng bằng sự phân bố về cường độ và hướng của cạnh, sau đó sẽ tiến hành hành phân lớp bằng ANNs bằng cách gán dữ liệu đầu vào (vectơ *n* chiều) vào lớp mong muốn.

Ưu điểm:

- Độ chính xác khi khảo sát thực tế cao (khoảng 94%).
- Thời gian nhận dạng nhanh (xấp xỉ 0.021 giây / ảnh).
- Nhận dạng được nhiều biển báo có hình dạng và màu sắc khác nhau.

Nhược điểm:

- Hạn chế khoảng cách nhận diện do ràng buộc tỉ lệ khung hình w/h < 1/3 và h/w < 1/7.
- Môi trường thực nghiệm chưa đa dạng, chỉ thực nghiệm trong môi trường ánh sáng bình thường.

1.3. Trần Hồng Việt, Đỗ Đình Tiến, Nguyễn Thị Trà, Trần Lâm Quân. Nhận dạng khuôn mặt sử dụng mạng nơron tích chập xếp chồng và mô hình FaceNet.

Hệ thống sử dụng mối tương quan giữa phát hiện và hiệu chỉnh để nâng cao hiệu suất trong một mạng nơ-ron tích chập xếp chồng (MTCNN) đồng thời kết hợp sử dụng framework FaceNet của Goole để tìm hiểu các ánh xạ từ hình ảnh khuôn mặt đến không gian Eculide, nơi khoảng cách tương ứng trực tiếp với độ đo độ tương tự khuôn mặt để trích xuất hiệu suất của các thuật toán đặc trưng khuôn mặt.

Tiền xử lý ảnh đầu vào: bao gồm phát hiện và cắt xén để lấy vùng ảnh chưa khuôn mặt, cải thiện chất lượng hình ảnh. MTCNN (Multi-task Cascaded Convolutional Networks) gồm 3 mạng CNN (Convolution, Relu, Max Pooling, Fully Connected Layers) xếp chồng và đồng thời hoạt động khi nhận dạng khuôn mặt. MTCNN hoạt động theo ba bước: P-Net

(Proposal Network) dự đoán các vùng chứa khuôn mặt trong ảnh, R-Net (Refine Network) sử dụng đầu ra của P-Net để loại bỏ các vùng không chứa khuôn mặt và mạng đầu ra là (Output Network).

Rút trích đặc trưng (FaceNet), thực hiện nhận dạng và phân lớp: FaceNet là một hệ thống nhúng cho việc nhận dạng và phân cụm khuôn mặt dựa trên việc nhúng mỗi ảnh vào không gian Euclide bằng cách sử dụng mạng CNN. Sau khi rút trích đặc trưng sẽ thực hiện nhận dạng bằng cách sử dụng mạng CNN VGG16 với kiến trúc bao gồm 13 lớp tích chập, 5 lớp max-pooling và 3 lớp kết nối đầy đủ, số lượng bộ lọc trong khối đầu tiên là 64, con số này được nhân đôi trong các khối tiếp theo đó cho đến khi đạt 512.

Ưu điểm:

- Nhận diện được mặt ở nhiều góc khác nhau, không cần nhìn thẳng.
- Độ chính xác cao và trích xuất được nhiều dữ liệu đặc trưng của khuôn mặt.

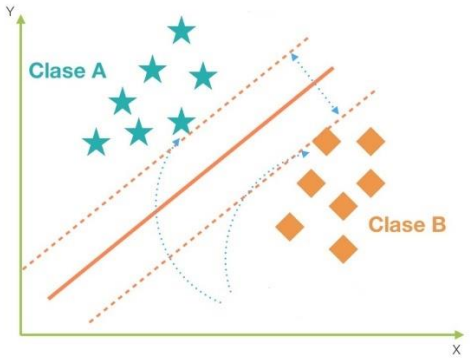
3. PHÁT BIỂU BÀI TOÁN

1.4. Bài toán

Bài toán chúng tôi đặt ra là phát hiện, nhận dạng nhiều biển báo giao thông không chồng lắp nhau dựa trên thuật toán CNN và SVM trong điều kiện ánh sáng bình thường. Sử dụng tập dữ liệu biển báo giao thông hình tròn của nước Đức (17909 ảnh) với 18 loại biển báo khác nhau.
Inputs: frame ảnh được chụp từ camera quay thời gian thực.
Outputs: hiển thị kết quả phân lớp (ảnh, thông tin biển báo).

1.5. Thuật toán

SVM (Support Vector Machine) là một giải thuật phân lớp (classification) nhằm xếp các mẫu dữ liệu hay các đối tượng vào một trong các lớp đã được định nghĩa trước. Các mẫu dữ liệu hay các đối tượng được xếp vào các lớp dựa vào giá trị của các thuộc tính (attributes) cho một mẫu dữ liệu hay đối tượng. Sau khi đã xếp tất cả các đối tượng đã biết trước vào các lớp tương ứng thì mỗi lớp được đặc trưng bởi tập các thuộc tính của các đối tượng chứa trong lớp đó. Quá trình phân lớp còn được gọi là quá trình gán nhã cho các tập dữ liệu. Hiểu đơn giản để phân lớp chỉ cần sử dụng một đường thẳng để phân tách các điểm nằm ở một bên là dương và các điểm còn lại là âm. Nếu có hai đường thẳng phân chia tốt thì chúng ta có thể phân tách khá xa hai tập dữ liệu (hình 1). SVM từng là một trong 10 giải thuật phân lớp hiệu quả, phổ biến trong cộng đồng khám phá tri thức và khai thác dữ liệu [Wu & Kumar, 2009], chính vì vậy chúng tôi đề xuất sử dụng mô hình SVM để nhận dạng các đối tượng biển báo giao thông.



Hình 1. Độ rộng lớn nhất được tính toán bởi một SVM tuyến tính.

Trong kỹ thuật SVM không gian dữ liệu nhập ban đầu sẽ được ánh xạ vào không gian đặc trưng và trong không gian đặc trưng này mặt siêu phẳng phân chia tối ưu sẽ được xác định. Ta có tập S gồm e các mẫu học:

$$S = \{(x1, y1), (x2, y2), (x3, y3) \dots (xe, ye)\} \subseteq (X \times Y)^{e}$$

Với một vectơ đầu vào *n* chiều $x_i \in R^n$ thuộc lớp I hoặc lớp II (tương ứng nhãn $y_i = 1$ đối với lớp I và $y_i = -1$ đối với lớp II). Một tập mẫu học được gọi là tầm thường nếu tất cả các nhãn là bằng nhau.

Đối với các dữ liệu phân chia tuyến tính, chúng ta có thể xác định đqợc siêu phẳng *f(x)* mà nó có thể chia tập dữ liệu. Khi đó, với mỗi siêu phẳng nhận được ta có: $f(x) \geq 0$ nếu đầu vào x thuộc lớp dương, và $f(x) < 0$ nếu *x* thuộc lớp âm trong đó *w* là vector pháp tuyến *n* chiều và *b* là giá trị ngưỡng.

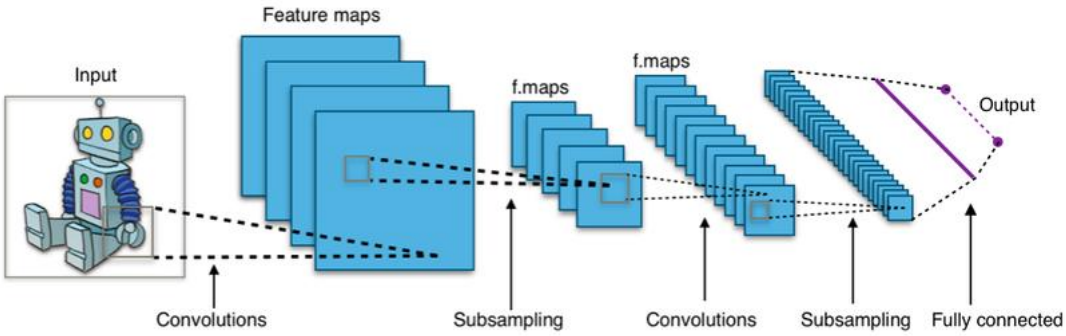
$$f(x) = w.x + b = \sum_{j=1}^{n} wixi + b$$

$$y_i f(x_i) = y_i (w.x_i + b) \geq 0, I = 1, \dots l$$

Vector pháp tuyến w xác định chiều của siêu phẳng f(x), còn giá trị ngưỡng b xác định khoảng cách giữa siêu phẳng và gốc.

Siêu phẳng có khoảng cách với dữ liệu gần nhất là lớn nhất (tức có biên lớn nhất) được gọi là siêu phẳng tối ưu.

CNN (Convolutional Neural Network) là một trong những mô hình Deep Learning tiến tiến cho phép xây dựng các hệ thống thông minh với độ chính xác cao. Sự ra đời của mạng CNN nhằm giúp giải quyết vấn đề giảm khối lượng tính toán bằng cách sử dụng các vùng tiếp nhận cục bộ, tập trọng số chia sẻ và phương pháp lấy tích chập để trích xuất thông tin thay cho các phương pháp cổ điển.



Hình 2. Mô hình CNN cơ bản

Lớp tích chập (convolutional layer): Lớp tích chập là một thành phần cốt lõi của mạng nơ-ron tích chập (CNN), sử dụng để trích xuất các thông tin đặc tính của hình ảnh (feature map) hỗ trợ cho quá trình "học" của mạng CNN. Phương thức hoạt động của lớp này được thực hiện thông qua quá trình trượt và lấy tích chập của bộ lọc (filter/kernel) trên toàn bộ ảnh. Kết quả đầu ra là đặc tính của ảnh tương ứng với bộ lọc đã sử dụng, với càng nhiều bộ lọc được sử dụng, chúng tôi sẽ thu được càng nhiều đặc tính của ảnh tương ứng.

Lớp lấy mẫu xuống (pooling layer): Lớp lấy mẫu xuống có tác dụng giảm kích thước của dữ liệu hình ảnh từ đó giúp cho mạng có thể học được các thông tin có tính chất khái quát hơn, đồng thời quá trình này giảm số lượng các thông số trong mạng. Các phương pháp lấy mẫu xuống thường được sử dụng là Max Pooling và Average Pooling.

Lớp dropout: Lớp dropout là một kỹ thuật được sử dụng để hạn chế hiện tượng overfitting (hiện tượng mạng nơ-ron quá bám sát vào tập dữ liệu huấn luyện và không đáp ứng được với các tập dữ liệu mới), thường gặp ở mạng CNN và giúp mô hình tính toán nhanh hơn. Dropout sử dụng phương pháp loại bỏ một số nơ-ron ngẫu nhiên trong mạng với một xác suất cho trước bằng cách thiết lập tất cả trọng số nơ-ron đó bằng 0, đồng nghĩa với các liên kết tới nơ-ron đó đều không có giá trị, khi đó mô hình sẽ phải cố gắng nhận dạng đúng trong khi thiếu thông tin từ các nơ-ron bị loại bỏ. Điều này sẽ giúp tăng tỉ lệ nhận dạng của mô hình nhưng không quá phụ thuộc vào dữ liệu huấn luyện.

Lớp kết nối đầy đủ (Fully-connected layer - FC): Đầu vào của lớp kết nối đầy đủ là đầu ra từ lớp lấy mẫu xuống hoặc lớp tích chập cuối cùng, nó được làm phẳng và sau đó được đưa vào lớp kết nối đầy đủ để chuyển tiếp. Lớp FC có nhiệm vụ tổng hợp thông tin đưa ra lớp quyết định (output) cho ra kết quả chính xác nhất.

HOG (Histrogram of oriented gradients) là một loại "feature descriptor", mục đích của "feature descriptor" là trừu tượng hóa các đối tượng bằng cách trích xuất ra những đặc trưng của đối tượng đó và bỏ đi những thông tin không hữu ích. Vì vậy chúng tôi sử dụng HOG cho mục đích trích đặc trưng biển báo phục vụ cho thao tác nhận dạng. Trích đặc trưng HOG trên ảnh gồm 4 bước:

Bước 1: Tính cường độ và hướng biến thiên tại mỗi pixels bằng công thức.

Cường độ: $|G| = \sqrt{I_x^2 + I^2 y}$

Hướng: $\theta = \frac{arctanI_x}{I_y}$

Bước 2: Chia ảnh đầu ra ở bước trên thành nhiều khối (block), mỗi khối có số ô bằng nhau, mỗi ô có số pixels bằng nhau. Các khối được xếp chồng lên nhau một ô. Số khối được tính bằng công thức. Trong đó, Wimage, Himage, Wblock, Hblock, Wcell, Hcell lần lượt là chiều rộng, chiều cao của ảnh, khối và ô.

$$n_{block/image} = \left(\frac{W_{image} - W_{block} - W_{cell}}{W_{cell}} + 1\right) * \left(\frac{H_{image} - H_{block} - H_{cell}}{H_{cell}} + 1\right)$$

Bước 3: Tính vectơ đặc trưng cho từng khối

Chia không gian hướng thành p bin (số chiều vectơ đặc trưng của ô).

Góc hướng nghiêng tại pixel (x, y) có độ lớn α (x, y) được rời rạc hóa vào một trong p bin.

Rời rạc hóa unsigned-HOG (p=9):

$$B\,(x, y) = round\,(\frac{p*a(x,y)}{\pi})mod\ p$$

Rời rạc hóa unsigned-HOG (p=18):

$$B\,(x, y) = round\,(\frac{p*a(x,y)}{2\pi})mod\ p$$

Giá trị bin được định lượng bởi tổng cường độ biến thiên của các pixels thuộc về bin đó. Nối các vectơ đặc trưng ô để được vectơ đặc trưng khối. Số chiều vectơ đặc trưng của khối tính theo công thức $size_{\text{feature/block}} = n_{cells} * size_{feature/cell}$. Trong đó, n_{cells} là số ô trong khối và $size_{feature/cell}$ là số chiều vectơ đặc trưng của ô bằng 9 (unsigned - HOG) hoặc 18 (signed - HOG).

Bước 4: Tính vectơ đặc trưng cho ảnh. Chuẩn hóa vectơ đặc trưng các khối bằng một trong các công thức:

L2-norm: $f = \frac{v}{\sqrt{v_2^2 + e^2}}$

L1-norm: $f = \frac{v}{(v_1 + e)}$

L1-sqrt: $f = \sqrt{\frac{v}{(v_1 + e)}}$

Trong các công thức trên, *v* là vectơ đặc trưng ban đầu của khối, v_k là k-norm của *v* (*k* = 1, 2), *e* là hằng số nhỏ.

Ghép các vectơ đặc trưng khối tạo nên ảnh để được đặc trưng R-HOG cho ảnh. Số chiều vectơ đặc trưng của ảnh tính theo công thức $size_{\text{feature/image}} = n_{blocs/image} * size_{\text{feature/block}}$, với $n_{blocs/image}$ là khối và $size_{\text{feature/block}}$ là số chiều vectơ đặc trưng mỗi khối.

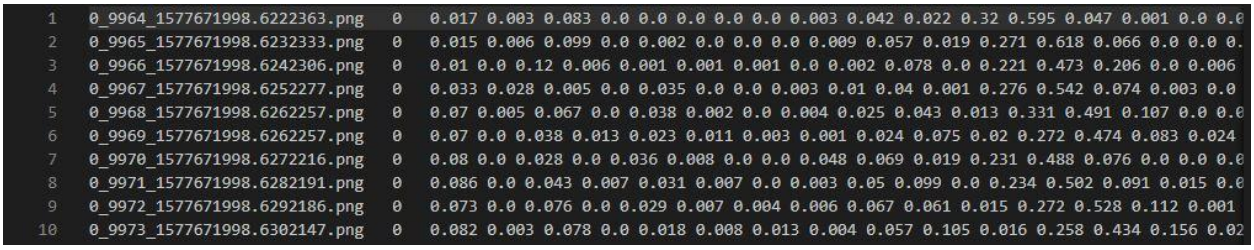
4. THỰC NGHIỆM

4.1. Dữ liệu

Chúng tôi sử dụng dữ liệu đầu vào là ảnh RGB có kích thước 32x32.

4.1.1. Tiền xử lý dữ liệu

Tiền xử lý dữ liệu sẽ bao gồm 3 giai đoạn:
Giai đoạn 1: Trích xuất khung ảnh từ camera để tiến hành phát hiện biển báo. Chuyển khung ảnh thành ảnh nhị phân, sau đó lọc nhiễu ảnh bằng hàm *cv2.medianBlur* để lộ rõ những vòng tròn (có thể là biển báo). Sử dụng hàm *cv2.HoughCircles* để tìm ra các biển báo hình tròn. Khi tìm thấy biển báo sẽ tiến hành cắt ảnh biển báo từ khung ảnh ban đầu.
Giai đoạn 2: Chuyển ảnh biển báo thành ảnh nhị phân, sau đó tiếp tục thực hiện cân bằng sáng và chuẩn hóa pixels từ 0 –> 255 thành 0 –> 1 để giảm kích thước của dữ liệu. Tất cả các quá trình trên đều sử dụng các phương thức do OpenCV hỗ trợ.
Giai đoạn 3: Từ dữ liệu thu được ở giai đoạn 1, tiếp tục thực hiện rút trích đặc trưng bằng cách sử dụng HOG. Kết quả các đặc trưng thu được sẽ lưu vào tập tin dưới dạng như hình 3 với cấu trúc là: tên tập tin ảnh / nhãn / các đặc trưng.



Hình 3: Các đặc trưng được trích ra từ ảnh sau khi sử dụng HOG

4.1.2. Phân chia dữ liệu

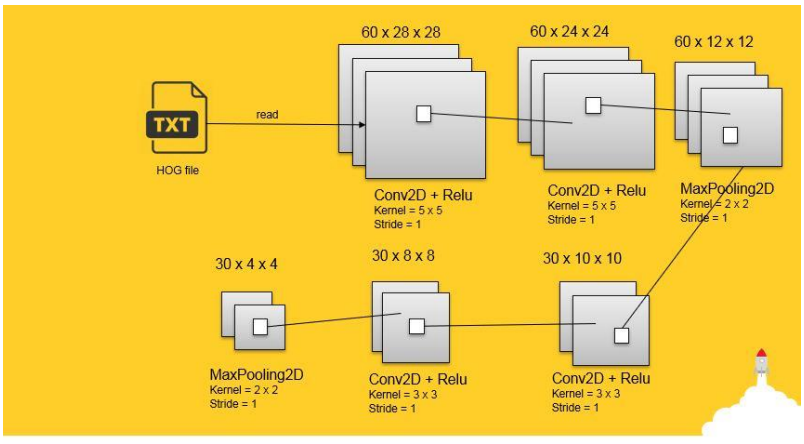
Dữ liệu được chia 80% (14327 ảnh) dùng để huấn luyện và 20% (3582 ảnh) dùng để kiểm tra.

Bảng 1. Tập dữ liệu

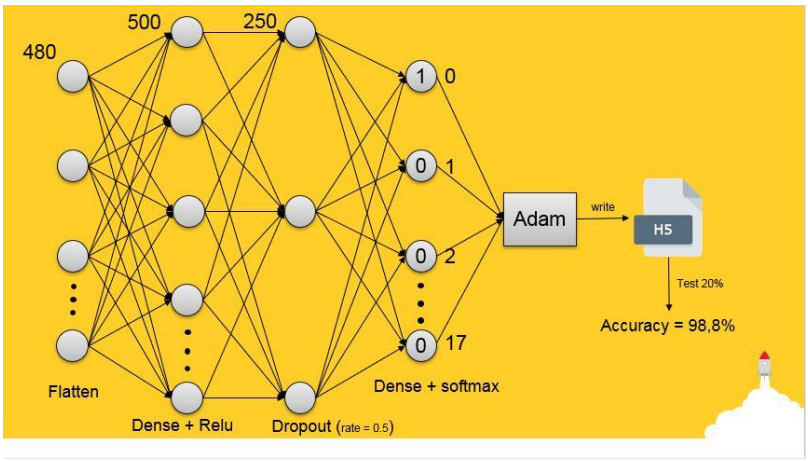
Mẫu	Số lượng huấn luyện	Số lượng kiểm tra	Nhãn (lớp)	Mẫu	Số lượng huấn luyện	Số lượng kiểm tra	Nhãn (lớp)
	144	36	0		792	198	9
	1584	396	1		497	102	10
	1608	402	2		288	72	11
	1008	252	3		864	216	12
	1416	354	4		264	66	13
	1320	330	5		144	36	14
	1032	58	6		1488	372	15
	1008	252	7		216	54	16
	432	108	8		240	60	17

4.2. Phương pháp:

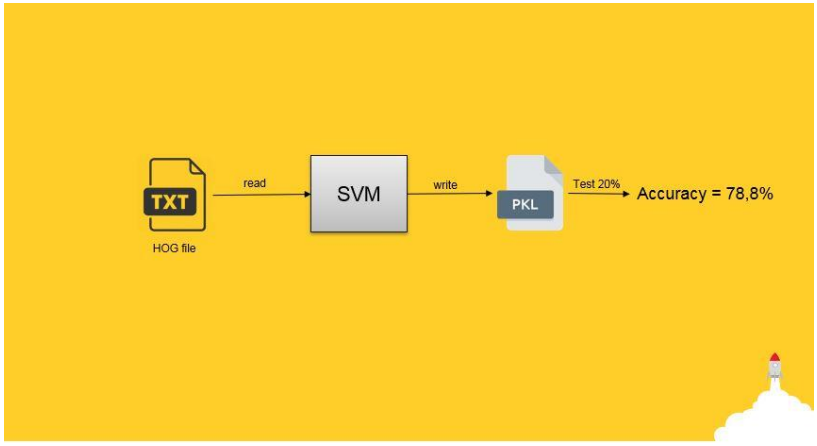
Dữ liệu sau khi tiền xử lý sẽ được lưu vào tập tin HOG, để tiến hành nhận dạng và phân lớp cần đưa dữ liệu đã được lưu ở tập tin HOG vào thuật toán.



Hình 4. Quy trình huấn luyện và đánh giá thuật toán CNN



Hình 5. Quy trình huấn luyện và đánh giá thuật toán CNN (tiếp theo)



Hình 6. Quy trình huấn luyện và đánh giá thuật toán SVM

4.3. Kết quả

Bảng 2. Kết quả huấn luyện

Thuật toán	Accuracy
CNN	98.8%
SVM	78.8%

Trong thực tế, chúng tôi đã tiến hành thực nghiệm với 18 loại biển báo được quay bằng điện thoại iphone 6 trong điều kiện ánh sáng ban ngày bình thường, điện thoại được đặt cố định với tốc độ quay chậm. Chúng tôi đã trích từng khung ảnh để xử lý và cho ra kết quả ở mức trung bình. Về độ chính xác SVM xấp xỉ 43%, CNN xấp xỉ 67%. Về tốc độ SVM nhận dạng một khung ảnh xấp xỉ 0.31 giây, CNN xấp xỉ 0.89 giây. Tuy nhiên, trường hợp số lượng biển báo nhiều dẫn đến tốc độ nhận dạng chậm, các biển báo bị chồng lắp cũng không thể nhận diện.

5. KẾT LUẬN

Mô hình huấn luyện máy học SVM và CNN sử dụng đặc trưng HOG trên các tập dữ liệu biển báo giao thông của nước Đức đạt kết quả ở mức trung bình. Kết quả thực nghiệm và nhận dạng một frame ảnh với thời gian khá nhanh nhưng độ chính xác không cao. Mô hình cũng cho thấy sự vượt trội của thuật toán CNN so với SVM khi CNN đạt độ chính xác xấp xỉ 67%, SVM chỉ có 43%. Với kết quả nghiên cứu này nếu muốn áp dụng vào thực tế cần phải cải thiện nhiều hơn về phần thuật toán nhận dạng cũng như xử lý ảnh đầu vào.

Trong tương lai, chúng tôi sẽ nghiên cứu cải tiến xử lý ảnh đầu vào để giải quyết trường hợp biển báo bị hư hỏng hoặc chụp trong điều kiện thiếu sáng. Tối ưu thuật cả hai thuật toán SVM và CNN để nâng độ chính xác và tốc độ nhận dạng của hệ thống. Ngoài ra, chúng tôi sẽ mở rộng phát triển tính năng tính khoảng cách từ camera đến biển báo.

TÀI LIỆU THAM KHẢO

Trương Quốc Bảo, Trương Hùng Chen, Trương Quốc Định, 2015. Phát hiện và nhận dạng biển báo giao thông đường bộ sử dụng đặc trưng HOG và mạng nơron nhân tạo. Tạp chí Khoa học Trường Đại học Cần Thơ, pp. 47-54, 2015.

Trần Hồng Việt, Đỗ Đình Tiến, Nguyễn Thị Trà, Trần Lâm Quân. Nhận dạng khuôn mặt sử dụng mạng nơron tích chập xếp chồng và mô hình FaceNet. Tạp chí khoa học và công nghệ, tập 57 – số 3, 6/2021.

Phạm Nguyên Khang, Trần Nguyễn Minh Thư, Đỗ Thanh Nghị, 2017. Điểm danh bằng mặt người với đặc trưng GIST và máy học véc-tơ hỗ trợ. FAIR, 2017.

Nikonorov, P. Yakimov, M. Petrov, Traffic sign detection on GPU using color shape regular expressions, VISIGRAPP IMTA-4, pp. 8, 2013.

P. Yakimov, Tracking traffic signs in video sequences based on a vehicle velocity [in Russian], Computer Optics. 39, 5, pp. 795-800, 2015.

Y. LeCun, P. Sermanet, Traffic Sign Recognition with Multi-Scale Convolutional Networks, Proceedings of International Joint Conference on Neural Networks (IJCNN'11), 2011.

M. Mathias, R. Timofte, R. Benenson, L. Gool, Traffic sign recognition - how far are we from the solution? Proceedings of IEEE International Joint Conference on Neural Networks, pp. 1-8, 2013.

H. Aghdam, E. Heravi, D. Puig, A practical approach for detection and classification of traffic signs using Convolutional Neural Networks, Robotics and Autonomous Systems. pp. 97-112, 2016.

H. Dean, and Jabir: “Real Time Detection and Recognition of Indian Traffic Signs using Matlab”, International Journal of Scientific & Engineering Research, Vol. 4, 684-690, 2013.

T. Surinwarangkoon, S. Nitsuwat, and E. J. Moore: “Traffic Sign Recognition System for Roadside Images in Poor Condition”, International Journal of Machine Learning and Computing, Vol. 3(1), pp. 121-126, 2013.

PHÂN CHIA CÔNG VIỆC

MSSV	Họ và tên	Nhiệm vụ	Chức vụ
18130054	Đoàn Lê Anh Duy	Phát biểu bài toán Thực nghiệm Kết luận	Nhóm trưởng
18130295	Lâm Hà Yến	Giới thiệu	
	Cao An Gia Lộc	Các công trình liên quan	
	Đoàn Quang Anh	Tóm tắt Tài liệu tham khảo	