**TRƯỜNG ĐẠI HỌC ĐÀ LẠT**

**KHOA CÔNG NGHỆ THÔNG TIN**

---🙘🙚---

****

**BÁO CÁO MÔN TRÍ TUỆ NHÂN TẠO**

**HỆ THỐNG PHÁT HIỆN BUỒN NGỦ KHI LÁI XE**

**Giáo viên: Nguyễn Thị Lương**

**Nhóm 8**

1914775 - Đinh Trọng Đạt

1911155 - Nguyễn Anh Nhật Huy

1910152 - Trương Quang Tuấn

Hệ Thống Phát Hiện Buồn Ngủ Khi Lái Xe

Nguyễn Thị Lương, Nguyễn Anh Nhật Huy – 1911155, Đinh Trọng Đạt - 1914775 , Trương Quang Tuấn – 1910152

*Lớp CTK43-PM, Khoa Công nghệ Thông tin, Trường Đại học Đà Lạt*

*Tóm tắt :* **Trong bài báo này, nhóm chúng tôi tập chung nghiên cứu, đề xuất một mô hình hiệu quả để giải quyết bài toán về việc phát hiện buồn ngủ khi lái xe, về công cụ, cách xây dựng công nghệ phát hiện buồn ngủ khi lái xe, tìm hiểu các thành phần cơ bản , phân tích và thiết kế công nghệ. Đồng thời chúng tôi cũng tìm hiểu về vai trò và cách thứ hoạt động của hệ thống phát hiện người lái xe buồn ngủ. Cuối cùng chúng tôi đánh giá và tìm ra các ưu nhược điểm của công nghệ từ đó đứa ra các giải pháp cải thiện và cũng như phát triển chúng theo nhiều hướng khác nhau. Phương pháp phát hiện buồn ngủ được đề xuất sử dụng ngôn ngữ lập trình Python, thư viện Keras, OpenCV và phương pháp CNN. Mô hình được đề xuất trên cơ sở mô hình mạng neural để trích lọc đặc trưng khuôn mặt. Hiệu quả của mô hình được kiểm nghiệm trên các tập cơ sở dữ liệu chuẩn. Các kết quả thực nghiệm cho thấy mô hình đề xuất đạt độ chính xác cao.**

***Từ khóa* – drowsiness detect, OpenCV, Eye-detection.**

I. GIỚI THIỆU

Hiện nay, vấn đề tham gia giao thông và an toàn giao là một trong những vấn đề khá quan trọng được nhiều chính phủ của các quốc gia trên thế giới quan tâm đến. Cơ quan an toàn giao thông đường cao tốc quốc gia (NHTSA) ước tính rằng lái xe buồn ngủ chiếm 91.000 vụ tai nạn giao thông, gây ra khoảng 50.000 thương tích và 800 trường hợp tử vong, theo báo cáo của cảnh sát năm 2017 [1]. Những số liệu thống kê trên về tai nạn giao thông cho thấy nguyên nhân gây tai nạn bởi sự thiếu tỉnh táo của tài xế chiếm một phần rất lớn, nguyên nhân dẫn đến có thể là do uống rượu bia, mệt mỏi do chạy xe trong thời gian dài… Những tài xế có thể cảm thấy buồn ngủ và có thể ngủ quên trong quá trình điều khiển xe, và tai nạn có thể xảy ra bất cứ lúc nào [2].

Buồn ngủ được coi là một sự chuyển đổi từ kích thích sang giấc ngủ [3]. Dấu hiệu đầu tiên của sự buồn ngủ là không có khả năng để giữ cho một đôi mắt mở. Việc đóng cửa một mắt thường xuyên khiến cho việc thực hiện nhiệm vụ của họ một cách hiệu quả. Tiếp theo, một đầu của một người có xu hướng bị lắc qua lại khi chúng buồn ngủ. Ngáp cũng là một dấu hiệu của sự buồn ngủ [3]. Lái xe khi một người bị thiếu ngủ tương tự như lái xe khi say rượu. Khi người ta cảm thấy buồn ngủ hơn, thời gian phản hồi tăng lên, khả năng thấy trước nguy hiểm giảm và độ bền của sự chú ý giảm. Lái xe không ngủ trong hơn 20 giờ tương đương với việc lái xe với nồng độ cồn trong máu là 0,08%, đó là giới hạn pháp lý ở Hoa Kỳ. Micros ngủ (thời gian ngắn vô thức vô thức) cũng có thể xảy ra ở một số người [2]. Trạng thái người lái xe có liên quan đến các yếu tố tâm lý, thể chất và tinh thần khác nhau [3]. Đặc biệt, mệt mỏi và đơn điệu được coi là giảm sự chú ý và mức độ kích thích, dẫn đến buồn ngủ [3]. Buồn ngủ có thể dẫn đến tai nạn bất cứ lúc nào trong ngày hay đêm. Những tai nạn như vậy do buồn ngủ có các đặc điểm sau [1] :

1. Thường xảy ra từ nửa đêm đến 6 giờ sáng hoặc vào cuối buổi chiều. Trong cả hai lần, nhịp sinh học, đồng hồ bên trong cơ thể kiểm soát giấc ngủ, bị giảm.
2. Trong nhiều trường hợp, một người lái xe duy nhất (không có hành khách) đã chạy khỏi đường với tốc độ cao mà không có dấu hiệu phanh.
3. Thường xảy ra trên đường địa phương và đường cao tốc.

Trong nghiên cứu này, chúng tôi đưa ra một giải pháp có thể làm cảnh báo những tài xế có biểu hiện buồn ngủ để giảm thiểu những rủi ro gây tai nạn giao thông trên đường - hệ thống cảnh báo buồn dựa trên thư viện OpenCV và Keras tập trung vào quan sát khả năng buồn ngủ của tài xế.

Mục tiêu của nhóm của nhóm đề ra khi làm đề tài này là hiểu được công cụ lập trình cũng như nâng cao tư duy về kỹ năng lập trình, giúp mọi người hiểu được quá trình xây dựng công nghệ và các phương pháp để đánh giá mô hình. Trên cơ sở từ những mục tiêu trên nhóm sẽ thu thập, nghiên cứu, tổng hợp các lý thuyết , kỹ thuật phương pháp để từ đó làm tiền đề để phát triểu ứng dụng đến hướng thực tế.

Bài báo này sẽ được viết theo cấu trúc như sau:

* Chương I: Giới thiệu đề tài.
* Chương II: Các nghiên cứu liên quan.
* Chương III: Phương pháp.
* Chương IV: Hệ thống phát hiện buồn ngủ khi lái xe (Drowsy Detection).
* Chương V: Cách thức hoạt động của công nghệ.
* Chương VI: Giới thiệu triển khai và đánh giá.
* Chương VII: Kết luật và hướng phát triển trong tương lai.
* Chương VIII: Tài liệu tham khảo về đề tài

II. CÁC NGHIÊN CỨU LIÊN QUAN

*1.* *Phát hiện buồn ngủ theo thời gian thực bằng tính năng Giám sát nháy mắt (Real Time Drowsiness Detection using Eye Blink Monitoring).*

Trong nghiên cứu này, một chiến lược đã được đề xuất phát hiện cơn buồn ngủ bằng cách sử dụng “Phát hiện trạng thái nháy mắt” (Eye state detection with Eye blinking strategy). Để bắt đầu, hình ảnh được chuyển đổi thành hình ảnh mờ và thuật toán phát hiện góc Harris được sử dụng để phát hiện các góc nằm ở đường cong của mắt và ở cả hai bên. Sau khi vẽ các điểm, một đường thẳng sẽ được vẽ giữa hai điểm trên và điểm giữa bằng cách tính đoạn thẳng và điểm giữa sẽ được nối với điểm dưới. Đối với mỗi hình ảnh, quy trình tương tự được thực hiện và khoảng cách ‘d’ từ tâm đến đáy được tính để xác định tình trạng của mắt. Cuối cùng, dựa trên khoảng cách ‘d’ được tính toán, ý định của trạng thái mắt được thực hiện. Trạng thái mắt sẽ được liệt kê là “nhắm” nếu d bằng 0 hoặc gần bằng 0, nếu không trạng thái mắt sẽ là “mở”. Họ có thể đã yêu cầu khoảng thời gian hoặc để biết liệu người đó có mệt hay không. Quá trình này mất 100-400 mili giây để hoàn thành nháy mắt của người bình thường.

*2. Phát hiện buồn ngủ trong thời lượng nháy mắt bằng thuật toán (Drowsiness Detection on Eye Blink Duration using Algorithm).*

Một chiến lược đã được đề xuất để nhận dạng buồn ngủ vào năm 2012. Webcam độ phân giải 640x480 được sử dụng để thực hiện nhận dạng nháy mắt liên tục. Mỗi lần nháy mắt được ước tính dựa trên giá trị trung bình được phân biệt với mỗi trường hợp. Làm sáng ở mỗi lần nhấp nháy với giá trị trung bình tiêu chuẩn mà khuôn khổ đó xem xét và cảnh báo được kích hoạt nếu giáo dục vượt qua khuyến khích này đối với một số đo cụ thể của các cạnh liên tiếp.

Chiến lược được đề xuất là hỗ trợ thuật toán máy vectơ. Những người sáng tạo đã ghi lại độ chính xác đến 99%. Đối với các điều kiện đang diễn ra, ở độ phân giải 640x480, khung chạy rất đáng kể. Trong tính toán này, khung công tác cần giữ dữ liệu về các cạnh trong quá khứ dựa trên thực tế là các ước tính về độ nheo mắt từ một số đo tổng hợp của các cạnh được sử dụng để hiển thị trên màn hình.

*3. Giám sát các tính năng trên khuôn mặt để phát hiện buồn ngủ thời gian thực (Facial Features Monitoring for Real Time Drowsiness Detection).*

Trong bài báo này, việc sử dụng các thuật toán sắc sảo trong xe đã phát triển rộng rãi trong những năm đang diễn ra. Để sàng lọc những gì khác, truyền trạng thái của xe và các khung lái xe sử dụng WSN. Để cải thiện bản chất của việc lái xe sắc sảo sử dụng các quy trình lập trình để điều khiển tốc độ động cơ, hướng dẫn, truyền, phanh và vv. Trong thời gian lựa chọn của người lái xe là yếu tố chính. Quan sát trạng thái của sự tồn tại và ngoại hình của các tài xế là một kỹ thuật khác để kiểm tra sự mệt mỏi của tài xế. Các hệ thống cảm biến từ xa không thể xử lý và truyền các dữ liệu này với độ chính xác thỏa đáng và đánh giá khá. Trong phương thức được cung cấp, video trước tiên được quay bằng camera sau đó nó được chia thành các phần trong phương pháp đề xuất. Mặt đầu tiên được phát hiện, và sau đó da được phân đoạn. Sau khi phân đoạn, hệ thống theo dõi mắt bằng thuật toán phát hiện cạnh, K có nghĩa là thuật toán được sử dụng để phát hiện ngáp. Sau đó, nó được đào tạo bằng máy vector hỗ trợ. Hệ thống cung cấp hoàn thành độ chính xác 94,58% tại bốn trường hợp thử nghiệm so với các kỹ thuật khác.

*4. Giám sát buồn ngủ dựa trên phát hiện ngáp (Driver Drowsiness Monitoring Based On Yawning Detection).*

Trong bài báo này, việc xác định hoặc sàng lọc người lái xe để có ý định liệu người lái xe có mệt mỏi hay không có ý nghĩa trong khuôn khổ này. Mệt mỏi được phân biệt trên ba: ước tính dựa trên sinh lý, xã hội và thực thi. Là ba chiến lược được đề xuất trong lĩnh vực này. Chiến lược đầu tiên tập trung vào nhận dạng mệt mỏi bằng cách nhận ra khuôn mặt và miệng để xác định ngáp. Kỹ thuật tiếp theo nhận ra khuôn mặt phụ thuộc vào sự phối hợp bố cục, sau đó sử dụng điều kiện bóng để xác định ngáp. Kỹ thuật cuối cùng sử dụng giả thuyết viola-Jones để nhận ra khuôn mặt và miệng. Xác định ngáp dẫn đến thăm dò ba chiến lược. Tìm kiếm khuôn mặt và miệng của tài xế phải chịu sự phân chia trong các điều kiện Khai sáng khác nhau trong cả chiến lược thứ nhất và thứ hai. Số lượng con người có bóng đặc biệt do đó dữ liệu bóng da khác nhau đã được sử dụng. Bóng môi là tâm điểm trong nhận dạng miệng.

*5. Phát hiện buồn ngủ dựa trên các tính năng của HOG và bộ phân loại SVM (Detection of Drowsiness based on HOG features and SVM classifiers).*

Bài viết này cung cấp phương pháp phát hiện sự buồn ngủ của hình ảnh có độ phân giải thấp. Phân loại Haar Cascade để theo dõi mắt được sử dụng để phát hiện sự buồn ngủ của trình điều khiển bằng cách kết hợp biểu đồ của gradient (HOG) định hướng nổi bật với phân loại máy vector (SVM) hỗ trợ để phát hiện chớp mắt. Perclos được tính toán khi phát hiện nhấp nháy mắt. Quy trình khác nhau của thuật toán bao gồm sử dụng camera để chụp các khung video, phát hiện khuôn mặt sau đó trích xuất điểm nổi bật, chiết vùng mắt và xác định mắt, xác định nhấp nháy, sau đó tính toán và phát hiện buồn ngủ. Có nhiều giai đoạn khác nhau trong buồn ngủ a) cực kỳ buồn ngủ, chống lại giấc ngủ b) buồn ngủ, một số nỗ lực để cảnh giác c) buồn ngủ, không gặp khó khăn gì trong việc cảnh giác. Nó mô tả một phương pháp phát hiện sự buồn ngủ của hình ảnh có độ phân giải thấp. Phân loại Haar Cascade để theo dõi mắt được sử dụng để phát hiện sự buồn ngủ của trình điều khiển bằng cách kết hợp biểu đồ của gradient (HOG) định hướng nổi bật với máy vectơ hỗ trợ. Giá trị perclos được tính sau khi phát hiện nhấp nháy. Nếu giá trị cạnh hơn 6 giây thì cá nhân bị buồn ngủ. Bằng cách tương phản kết quả với các quan sát người đánh giá của con người, chương trình đã được mở rộng. Nó sản xuất 91,6%.

*6. Hệ thống EEG có thể đeo dựa trên smartwatch để phát hiện buồn ngủ của người lái (Smartwatch based Wearable EEG System for Driver Drowsiness Detection).*

Lý do tử vong trong va chạm tự động trên khắp thế giới là buồn ngủ của người lái xe. Để nhận ra sự mệt mỏi của tài xế. Nhiều dấu hiệu sinh lý đã được đề xuất. Điện não đồ (Electroencephalographic) là một trong những dấu hiệu này, trong đó phản ánh các bài tập tâm trí, tất cả đều được xác định hợp pháp hơn với sự lười biếng. Những mô hình chỉ thước đo các dấu hiệu riêng biệt là một trong những trở ngại của những điều này kỳ thi. Ngoài ra, dọc theo những dòng này đã không đi vào xem xét đánh giá mức độ nghiêm trọng tương đối của người lái uể oải. Có ba giai đoạn buồn ngủ bao gồm. Cảnh báo, cảnh báo sớm, buồn ngủ. Hỗ trợ máy vector dựa trên. Mô hình xác suất trở lại (SVMPPM) được đề xuất cho điều này. Kiểm tra DDD ((Detecting Driver Drowsiness) tập trung vào việc thay đổi mức độ chậm chạp thành bất kỳ. Ước tính 0 ~ 1 thay vì tên riêng biệt. Xây dựng mô hình này hai mươi đối tượng được sử dụng. Để xây dựng mô hình mười lăm đối tượng và để kiểm tra mô hình năm đối tượng được sử dụng. Nó đưa ra độ chính xác của cảnh báo: 91,25% cảnh báo sớm: 83,78% buồn ngủ: 91,92%.

III. Các phương pháp.

*A. Phát hiện đối tượng trên khung hình*

Như đã đề cập ở trên, bản chất của việc phát hiện đối tượng là quá trình tìm kiếm và định vị đối tượng trên khung ảnh bất kỳ. Phương pháp phát hiện khuôn mặt ở đây được đề xuất sử dụng các đặc trưng HOG (Histográm of Oriented Gradients) và bộ phân lớp tuyến tính SVM (Support Vector Machines).

Gom nhóm đặc trưng tại mỗi cell

Tính đạo hàm gradient

frame

Frame

Có mặt

Phân lớp tuyến tính SVM

Tính đặc trưng trên các cửa sổ detector

Tính đặc trung cho khối, chuẩn hóa

Không mặt

***Hình 6****. Phương pháp phát hiện khuôn mặt*

Ý tưởng chính của đặc trưng HOG là hình dạng và trạng thái của vật thể được đặc trưng bởi sự phân bổ về gradient và hướng của cạnh. Đặc trưng này được phát triển dựa trên các đặc trưng SIFT (Scale-Invariant Feature Transform), đặc trưng HOG được tính trên cả một vùng, Do sự biến thiên màu sắc trong các vùng khác nhau nên mỗi vùng sẽ cho ta một vector đặc trưng của nó. Vì vậy để có được đặc trưng của toàn bộ cửa sổ ta phải kết hợp nhiều vùng liên tiếp lại với nhau. Các bước cơ bản trong quy trình phát hiện khuôn mặt/đối tượng trên các khung hình được mô tả cụ thể trên *Hình 6*.

Đầu vào của thuật toán là một frame ảnh bất kỳ thu được từ bước phân đoạn video. Bước xử lý đầu tiên sẽ tiến hành chuyển đổi ảnh trong không gian RGB sang ảnh đa cấp xám, sau đó tiến hành cân bằng histogram trên ảnh gray scale để giảm sự nhạy cảm với nguồn sáng.

Việc lưu trữ chính xác từng giá trị gốc của từng điểm ảnh (x,y) tốn nhiều chi phí và không mang lại nhiều kết quả, do vậy ta sẽ chia không gian gốc ra thành các bin. Việc phân chia bin càng nhỏ sẽ càng làm tăng độ chính xác, các kết quả thử nghiệm cho thấy kích thước bin khoảng 200 cho kết quả tốt nhất đối với việc phát hiện khuôn mặt người. Do đó với không gian hướng biến thiên trong miền từ 0° - 180° sẽ được chia thành 9 bi như sau: [0° -20°], [21° - 40°], [41° - 60°], [61° - 80°], [81° - 100°], [101° - 120°], [121° - 140°], [141° - 160°], [161° - 180°]. Ứng với mỗi bin trên, tiến hành thống kê biên độ (magnitude) tại từng vị trí, với mỗi bin, tại vị trí (x,y) nếu góc (orientation) thuộc về bin đó thì giá trị của bin đó tại vị trí (x,y) bằng 0. Bước tiếp theo tiến hành tính toán vector đặc trưng cho từng cell (mỗi cell thường được chọn với kích thước 8x8 px). Vector đặc trưng của mỗi cell sẽ gồm 9 thành phần tương ứng với 9 bin và giá trị tại thành phần I bằng tổng giá trị của các điểm trong bin I mà có toạ độ nằm trong cell đó. Tiếp theo, tính toán vector đặc trưng của từng cell trong block lại với nhau.

Với giả thiết mỗi cell có kích thước 8x8 px, mỗi block có kích thước 2x2 cells, không gian hướng biến thiên xét trong miền từ 0° - 180° và được chia thành 9 bin thì số đặc trưng trong mỗi khối sẽ được tính bằng 4x9=36 thành phần. Từ đó, tiến hành tính toán vector đặc trưng của các cửa sổ trên toàn bộ ảnh đầu vao. Trong đó, một cửa sổ được tạo bởi các khối xếp gối nhau – overlapping. Đặc trưng của một cửa sổ sẽ được tính bằng cách ghép các vector đặc trưng của từng block tạo lên cửa sổ đó.

Ở bước xử lý cuối cùng, toàn bộ vector đặc trưng thu được trên mỗi cửa sổ sẽ được sử dụng làm đầu vào của bộ phận lớp tuyến tính SVM. Bộ phân lớp có nhiệm vụ xác định lớp mẫu (nhắm mắt hay mở mắt) đối với mỗi ảnh đầu vào dựa trên các tri thức mà thuật toán đã huấn luyện.

*B. Phát hiện buồn ngủ bằng phương pháp CNN.*

Công đoạn phát hiện nhận dạng thường gồm 2 bước xử lý chính là trích cọn đặc trưng và phân lớp khuôn mặt. Phương pháp trích chọn đặc trưng ở đây được đề xuất sử dụng các lớp mạng neural tích chập (CNN – Convolutional Neural Network), là một tập hợp con của học máy và chúng là trung tâm của các thuật toán học sâu. CHúng bao gồm các lớp nút, chứa một lớp đầu vào, một hoặc nhiều lớp ẩn và một lớp đầu ra. Mỗi nút kết nối với một nút khác và có trọng lượng và ngưỡng liên quan. Nếu đầu ra ca bất kỳ nút riêng lẻ nào vượt quá giá trị ngưỡng được chỉ điịnh, nút đó sẽ được kích hoạt, gửi dữ liệu đến lớp tiếp theo của mạng. Mạng neural tích chập thường được sử dụng để phân loại các tác vụ thị giác máy tính. Trước CNN, các phương pháp trích xuất tính năng thủ công, tổng thời gian đã được sử dụng để xác định các đối tượng trong hình ảnh. Tuy nhiên, các mạng neural tích chập hiện cung cấp một cách tiếp cận có thể nở rộng hơn đối với các tác vụ phân loại hình ảnh và nhận dạng đối tượng, tận dụng các nguyên tắc từ đại số tuyến tính, đặc biệt là phép nhân ma trận, để xác định các mẫu trong hình ảnh. Điều đó nói rằng, chúng có thể đòi hỏi tính toán, yêu cầu các đơn vị xử lý đồ hoạ (GPU) để đào tạo các mô hình.

Diagram

Description automatically generated

***Hình 7****. Convolutional neural networks (CNN)*

Đây là mô hình có khả năng học từ một tập mẫu cho trước, nhằm tự động phát hiện các đặc trưng quan trọng nhất để nhận dạng đối tượng (ở đây là khuôn mặt đeo khẩu trang). Ý tưởng chính của hướng tiếp cận này dựa trên việc học một không gian Euclidean nhúng trong mỗi ảnh sử dụng một cấu hình mạng neural tích chập. Mạng được huấn luyện sao cho khoảng cách 1,2 bình phương trong không gian nhúng là tương ứng trực tiếp vói độ tương tự của khuôn mặt.

A picture containing chart

Description automatically generated

***Hình 8****. Cách thức hoạt động của CNN*

CNN được phân biệt với các mạng neural khác bởi hiệu suất vượt trội của chúng với đầu vào tín hiệu hình ảnh, lời nói hoặc âm thanh. Chúng có ba loại chính, đó là:

* Convolutional Layer.
* Pooling Layer.
* Fully-Connected Layer.

Convolutional layer là lớp đầu tiên của mạng tích chập. Mặc dù các lớp tích chập có thể được theo sau bởi các lớp tích chập bổ sung hoặc các lóp gộp chung, lớp được kết nối đầy đủ là lớp cuối cùng. Với mỗi lớp, CNN tăng độ phức tạp của nó, xác định các phần lớn hơn của hình ảnh. Các lớp trước đó tập trung vào các tính năng đơn giản, chẳng hạn như màu sắc và cạnh. Khi dữ liệu hình ảnh tiến triển qua cá lớp của CNN, nó bắt đầu nhận ra các phần tử hoặc hình dạng lớn hơn của đối tượng cho đến khi cuối cùng nó xác định đối tượng dự định.

Convolutional layer là khối xây dựng cốt lõi của CNN và đó là nơi xẩy ra phần lớn tính toán. Nó yêu cầu một vài thành phần, đó là dữ liệu đầu vào, bộ lọc và bản đồ tính năng. Giả sử rằng đầu vào sẽ là một hình ảnh màu, được tạo thành từ một ma trận các điểm ảnh trong 3D. Điều này có nghĩa là đầu vào sẽ có ba chiều-chiều cao, chiều rộng và độ sâu-tương ứng với RGB trong ảnh. Chúng tôi cũng có một máy dò tính năng, còn được gọi là hạt nhân hoặc bộ lọc, sẽ di chuyển qua các trường tiếp nhận của hình ảnh, kiểm tra xem tính năng này có hiện diện không. Quá trình này được gọi là một Convolutional.

Máy dò chức năng là một mảng trọng lượng hai chiều (2-D), đại diện cho một phần của hình ảnh. Mặc dù chúng có thể khác nhau về kích thước, kích thước bộ lọc thường là ma trận 3x3; điều này cũng xác định kích thước của trường tiếp nhận. Bộ lọc sau đó được áp dụng cho một khu vực của hình ảnh và một sản phẩm dấu chấm được tính toán giữa các điểm ảnh đầu vào và bộ lọc. Sản phẩm này sau đó được đưa cho một mảng đầu ra. Sau đó, bộ lọc thay đổi theo một sải bước, lặp lại quá trình cho đến khi hạt nhân quét qua toàn bộ hình ảnh. Đầu ra cuối cùng từ loạt sản phẩm dấu chấm từ đầu vào và bộ lọc được gọi là bản đồ tính năng, bản đồ kích hoạt hoặc tính năng xoay

A picture containing diagram

Description automatically generated

***Hình 9****.Mô hình CNN*

Diagram

Description automatically generated

***Hình 10****. Các lớp trong mô hình CNN*

Như bạn có thể thấy trong *Hình 9, Hình 10*, mỗi giá trị đầu ra trong bản đồ tính năng không phải kết nối với từng giá trị pixel trong hình ảnh đầu vào. Nó chỉ cần kết nối với trường tiếp nhận, nơi bộ lọc đang được áp dụng. Vì mảng đầu ra không cần ánh xạ trực tiếp đến từng giá trị đầu vào, các Convolutional layer thường được gọi là các lớp “được kết nối một phần”. Tuy nhiên, đặc điểm này cũng có thể được mô tả là kết nối cục bộ.

Lưu ý rằng trọng lượng trong máy dò tính năng vẫn cố định khi nó di chuyển trên hình ảnh, còn được gọi là chia sẻ tham số. Một số tham số, như giá trị trọng lượng, điều chỉnh trong quá trình đào tạo thông qua quá trình backpropagation và gradient descent. Ty nhiên, có ba siêu cận điểm ảnh hưởng đến kích thước âm lượng của đầu ra cần được đặt trước khi đào tạo mạng thần kinh bắt đầu. Chúng bao gồm:

1. Số lượng bộ lọc ảnh ảnh hưởng đến độ sâu của đầu ra. Ví dụ: ba bộ lọc riêng biệt sẽ mang lại ba bản đồ tính năng khác nhau, tạo ra độ sâu ba.
2. Sải (Stride) là khoảng cách, hoặc số điểm ảnh, mà hạt nhân di chuyển qua ma trận đầu vào. Trong khi các giá trị sải của hai hoặc lớn hơn là rất hiếm, một sải bước lớn hơn mang lại chất lượng thấp hơn.
3. Zero-padding thường được sử dụng khi các bộ lọc không phù hợp với hình ảnh đầu vào. Điều này đặt tất cả các phần tử nằm ngoài ma trận đầu vào về 0, tạo ra đầu ra lớn hơn hoặc có kích thước bằng nhau. Có 3 loại padding: Valid padding, Same padding, Full padding.

Cấu trúc của CNN có thể trở nên phân cấp vì các lớp sau có thể nhìn thấy các điểm ảnh trong các trường tiếp nhận của các lớp trước đó. Cuối cùng, Convolutional layer chuyển đổi hình ảnh thành các giá trị số, cho phép mạng neural diễn giải và trích xuất các mẫu có liên quan.

Diagram

Description automatically generated

***Hình 11****. Max Pooling layer.*

Pooling layer tiến hành giảm tính chiều hướng, giảm số lượng tham số đầu vào. Tương tự nhue Convolutional layer, hoạt động gộp chung quét một bộ lọc trên toàn bộ đầu vào, nhưng sự khác biệt là bộ lọc này không có bất kyù trọng lượng nào. Thay vào đó, hạt nhân áp dụng hàm tổng hợp cho các giá trị trong trường tiếp nhận, điền mảng đầu ra. Mặc dù rất nhiều thông tin bị mất trong Pooling layer, nhưng nó cũng có một số lợi ích cho CNN. Chúng giúp giảm độ phức tạp, cải thiện hiệu quả và hạn chế nguy cơ quá tải.

Fully-Connected layer, như đã đề cập trước đó, các giá trị pixel của hình ảnh đầu vào không được kết nối trực tiếp đến với lớp đầu ra trong các lớp được kết nối một phần. Tuy nhiên, trong layer được kết nối đầy đủ, mỗi nút trong layer output kết nối trực tiếp với một nút trong layer trước đó.

Lớp này thực hiện nhiệm vụ phân loại dựa trên các tính năng được trích xuất thông qua các lớp trước đó và các bộ lọc khác nhau của chúng. Trong khi các convolutional và pooling có xu hướng sử dụng các hàm ReLu, các lớp FC thường tận dụng chức năng kích hoạt softmax để phân loại đầu vào một các thích hợp, tạo ra xác suất từ 0 đến 1.

Đi sâu vào mô hình phát hiện buồn ngủ, chúng tôi sử dụng phương pháp CNN để training dữ liệu cho máy. Dữ liệu được train cho máy gồm Những file hình ảnh vùng mắt có trạng thái đang nhắm hoặc mở. CNN sẽ tạo ra các lớp để phân tích, học tập các đặc trưng quan trọng nhất của file ảnh, để tiến hành quá trình nhận dạng.

Mạng neural được training một cách trực tiếp để đầu ra của nó trở thành một vector đặc trưng 128 chiều, sử dụng hàm chi phí bộ ba (tripletbased loss function). Một bộ ba (triplet) được định nghĩa bao gồm các vùng mắt của nhiều người khác nhau. Mục tiêu của hàm là phân tách các vùng mắt ra khỏi khuôn mặt negative sử dụng một lề khoảng cách – distance margin. Từ các độ đo thu được, thuật toán sẽ ước lượng giá trị của hàm chi phí dựa trên việc so sánh khoảng cách giữa các tập đặc trưng được sinh ra từ các ảnh khuôn mặt đẹo khẩu trang khác nhau. Các giá trị ước lượng của hàm chi phí sau khi tính sẽ được lan truyền ngược từ lớp cuối cùng đến lớp đầu tiên của mạng để tính chi phí trọng số trên các lớp mạng. Quá trình tính toán, ước lượng và cập nhật trọng số của mạng được lặp đi lặp lại liên tục cho đến khi giá trị của hàm chi phí thỏa mãn điều kiện đã cho. Lặp lại các bước trên đối với toàn bộ tập dữ liệu huấn luyện cho đến khi thuật toán huấn luyện mạng hội tụ.

IV. HỆ THỐNG PHÁT HIỆN BUỒN NGỦ KHI LÁI XE

*1. Các khái niệm.*

Xử lý ảnh là một trong những mảng quan trọng nhất trong kỹ thuật thị giác máy tính, là tiền đề cho nhiều nghiên cứu thuộc lĩnh vực này. Hai nhiệm vụ cơ bản của quá trình xử lý ảnh là nâng cao chất lượng thông tin hình ảnh và xử lý số liệu cung cấp cho các quá trình khác trong đó có việc ứng dụng thị giác vào điều khiển.

Quá trình bắt đầu từ việc thu nhận ảnh nguồn (từ các thiết bị thu nhận ảnh dạng số hoặc tương tự) gửi đến máy tính. Dữ liệu ảnh được lưu trữ ở định dạng phù hợp với quá trình xử lý. Người lập trình sẽ tác động các thuật toán tương ứng lên dữ liệu ảnh nhằm thay đổi cấu trúc ảnh phù hợp với các ứng dụng khác nhau.

Quá trình xử lý nhận dạng ảnh được xem như là quá trình thao tác ảnh đầu vào nhằm cho ra kết quả mong muốn. Kết quả đầu ra của một quá trình xử lý ảnh có thể là một ảnh “đã được xử lý” hoặc một kết luận. Rất nhiều hướng tiếp cận đã được đề xuất để giải quyết bài toán này [9] [5] [4] [2]. Nhìn chung, quy trình giải quyết bài toán thường bao gồm các công đoạn cơ bản như: *(i)* Thu thập hình ảnh; *(ii)* Tiền xử lý, tang cường chất lượng hình ảnh; *(iii)* Phát hiện, căn chỉnh, cắt ảnh đối tượng; *(iv)* Nhận dạng (trích chọn đặc trưng và phân lớp) đối tượng.

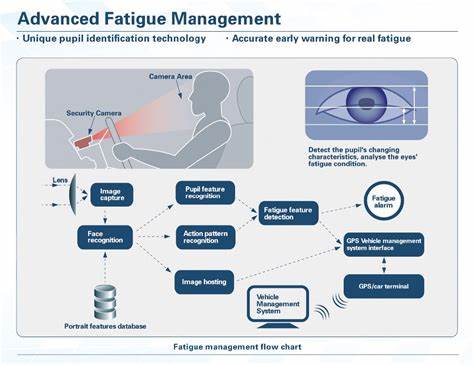
Bài toán phát hiện và phân loại đối tượng là bài toán đã được nghiên cứu từ những năm 70 của thế kỷ XX. Tuy nhiên, đây là một bài toán khó, nên những nghiên cứu hiện tại vẫn chưa đạt được những kết quả mong muốn. Chính vì thế, vấn đề này vẫn đang được nhiều nhóm trên thế giới quan tâm nghiên cứu. Khó khăn của bài toán phát hiện và phân loại đối tượng có thể kể đến như sau:

* Tư thế chụp, góc chụp: Ảnh chụp khuôn mặt có thể thay đổi rất nhiều bởi vì góc chụp giữa camera và khuôn mặt. Chẳng hạn như chụp thẳng, chụp chéo bên trái 45°, chụp từ trên xuống, chụp từ dưới lên, v.v… Với các tư thế khác nhau, cá thành phần trên khuôn mặt như mắt, mũi, miệng có thể bị khuất một phần hoặc thậm chí khuất hết, gây khó khăn cho những bài toán phát hiện đối tượng như mắt, mũi , miệng.
* Sự xuất hiện hoặc thiếu một số thành phần của khuôn mặt: Các đặc trưng như mắt kính, nón, v.v… có thể xuất hiện hoặc không. Vấn đề này làm cho bài toán càng trở nên khó hơn rất nhiều.
* Sự biểu cảm của khuôn mặt: Biểu cảm của khuôn mặt con người có thể làm ảnh hưởng đáng kể lên các thông số của đối tượng. Chẳng hạn, cùng một khuôn mặt một người, nhưng có thể sẽ rất khác khi họ cười hoặc sợ hãi, v.v…
* Sự che khuất: Khuôn mặt có thể bị che khuất bởi các đối tượng khác hoặc các đối tượng khác.
* Hướng của ảnh (pose variations): Các ảnh khuôn mặt có thể biến đối rất nhiều với các góc quay khác nhau của trục camera. Chẳng hạn chụp với trục máy ảnh nghiêng làm cho khuôn mặt bị nghiêng so với trục của ảnh.
* Điều kiện của ảnh: Ảnh được chụp trong các điều kiện khác nhau về: chiếu sáng, về tính chất camera (máy kỹ thuật số, máy hồng ngoại, v.v…), ảnh có chất lượng thấp ảnh hưởng rất nhiều đến chất lượng ảnh khuôn mặt.
* Aging condition: Việc phát hiện đối tượng thay đổi theo thời gian còn là một vấn đề khó khăn, ngay cả đối với khả năng nhận dạng của con người.
* Các hệ thống cực lớn: Các cơ sở dữ liệu ảnh mắt được test bởi các nhà nghiên cứu còn khá nhỏ (vài trăm tới vài chục nghìn ảnh mắt), tuy nhiên trên thực tế, các cơ sở dữ liệu có thể rất lớn, ví dụ cơ sở dữ liệu ảnh mặt của cảnh sát trong một đất nước có thể chứa từ hàng triệu tới hơn một tỉ ảnh.

Phát hiện khuôn mặt, dò tìm, định vị những vùng (vị trí) có thể là khuôn mặt xuất hiện trong ảnh hoặc các frame video. Các vùng này sẽ được tách riêng để xử lý. Phân đoạn khuôn mặt sẽ xác định vị trí mắt mũi, miệng và các thành phần khác của khuôn mặt và chuyền kết quả này cho bước trích chọn đặc trưng. Ở bước trích chọn đặc trưng, bằng một phương pháp trích chọn đặc điểm nào đó (mẫu nhị phân cục bộ-Local Binary Pattern-LBP, Gabor wavelets,…) sẽ được sử dụng với ảnh đối tượng để trích xuất các thông tin đặc trưng. Kết quả của mỗi ảnh sẽ được biểu diễn dưới dạng một vector đặc trưng. Những vector đặc trưng này sẽ là dữ liệu đầu vào cho một mô hình đã được huấn luyện trước để phát hiện và phân loại đối tượng (Object detection).

Hệ thống phát hiện buồn ngủ cho người lái xe (DDS) là một phần của công nghệ an toàn xe hoạt động dựa trên thuật toán phát hiện những thay đổi trong hành vi lái xe của người lái, chẳng hạn như chuyển động bánh xe thất thường, chệch làn đường, khó giữ mắt và liên tục ngáp, v.v. [9]

Một số hệ thống cảnh báo người lái xe nghỉ giải lao bằng cảnh báo âm thanh, trong khi một số hệ thống hiển thị biểu tượng cà phê và một số xe ô tô thậm chí còn rung ghế lái. [9]



***Hình 4*** *: Sơ đồ về hệ thống phát hiện buồn ngủ khi lái xe*

*2.Tầm quan trọng và tính thiết thực.*

Những ảnh hưởng của lái xe mệt mỏi có thể gây nguy hiểm cho mọi người trên đường. Người lái xe buồn ngủ sẽ mất thời gian để tập trung, phản ứng chậm và không thể phán đoán được tốc độ và khoảng cách. [9]

Một người lái xe buồn ngủ không phải lúc nào cũng là người chưa ngủ đủ giấc. Do đó, điều quan trọng là phải phát triển một công cụ để cảnh báo những người lái xe mệt mỏi về nguy hiểm sắp xảy ra. Bạn phải có đủ bộ dữ liệu để đào tạo mô hình học máy và nhận dạng khuôn mặt để thực hiện điều này. [9]

Để đào tạo chính xác mô hình DDS, bạn cần một bộ sưu tập toàn diện các bộ dữ liệu đào tạo (chứa cả hình ảnh buồn ngủ và không buồn ngủ của người) có thể giúp định vị các điểm mốc trên khuôn mặt trên hình ảnh. Phương pháp này giúp hệ thống xác định các đặc điểm khuôn mặt của người lái xe trong các tình huống thời gian thực.

Ngoài ra, vì hệ thống đặc biệt quan tâm đến mắt, tọa độ được hiển thị cho mắt, điều này sẽ giúp phát hiện các giá trị nhấp nháy và mở mắt.

Các tập dữ liệu có chứa hình ảnh có thể giúp hệ thống nhận ra hành động ngáp cũng nên được đưa vào. Ngoài khả năng phát hiện nháy mắt, ngáp cũng là một thông số quan trọng mà hệ thống phải tìm hiểu để đưa ra cảnh báo cho người lái. Một mô hình học máy có thể được xây dựng bằng cách sử dụng các bộ dữ liệu được gắn nhãn chính xác và các phương pháp học sâu.

Dữ liệu cho một hệ thống phát hiện buồn ngủ được chia làm 3 tập dữ liệu: tập dữ liệu huấn luyện (training set), tập dữ liệu tham chiếu (reference set) và tập dữ liệu để nhận dạng, kiểm tra độ chính xác (test set). Trong nhiều hệ thống, tập dữ liệu huấn luyện trùng với tập dữ liệu tham chiếu. Tập huấn luyện gồm các ảnh được dùng để huất luyện, thông thường, tập này được dùng để sinh ra một không gian con là một ma trận. Tập tham chiếu gồm các ảnh đã biết danh tính được chiếu vào không gian con ở bước huấn luyện (Training). Bước huấn luyện nhằm 2 mục đích: giảm số chiều (dimension reduction) của các vector đặc điểm (feature vector) vì các vector này thường có độ dài khá lớn nên nếu để nguyên thì việc tính toán sẽ rất rất lâu, thứ hai, làm tăng tính phân biệt (discriminative) giữa các ảnh khác lớp, ngoài ra có thể làm giảm tính phân biệt giữa các ảnh thuộc về một lớp, ví dụ như Linear Discriminant Analysis LDA – còn gọi là Fisher Linear Discriminant Analysis là một phương pháp làm việc với tập training mà mỗi đối tượng có nhiều ảnh mặt ở các điều kiện khác nhau. Sau khi thực hiện chiếu tập reference vào không gian con, hệ thống lưu lại kết quả là một ma trận với mỗi cột của ma trận là một vector tương ứng với ảnh để thực hiện phát hiện, nhận dạng hay phân lớp. Phân lớp được thực hiện với tập các ảnh, sau khi tiền xử lý xong, mỗi ảnh sẽ được áp dụng phương pháp trích chọn đặc điểm và được chiếu vào không gian con. Tioeép đến việc phân lớp sẽ dựa trên phương pháp k-NN, định danh của một ảnh cần xác định sẽ được gán là định danh của ảnh, có khoảng cách gần với nó nhất. Ở đây cần lưu ý, mỗi ảnh là một vector nên có thể dùng khái niệm hàm khoảng cách giữa hai vector để đo sự khác biệt giữa các ảnh.

Sự cần thiết của một chính xác Lái xe buồn ngủ Hệ thống phát hiện tiếp tục phát triển. Các doanh nghiệp đang tìm kiếm bộ dữ liệu đào tạo có độ tin cậy cao có thể được sử dụng để đào tạo các mô hình ML của họ.

Khi cần độ tin cậy và sự đa dạng trong bộ dữ liệu, nhiều nhà cung cấp công nghệ hàng đầu thích Shaip hơn. Shaip đã là công cụ trong việc phát triển các mô hình DDS cao cấp với bộ dữ liệu đa dạng, ghi nhãn hình ảnh chất lượng cao và chú thích. Có một ứng dụng DDS phá vỡ đường dẫn trong tâm trí? Kết nối với Shaip và khám phá các bộ dữ liệu đào tạo đa dạng với giá cả cạnh tranh. [9]

*3.Ưu nhược điểm*

Có rất nhiều ưu điểm của DDS, và lợi ích đầu tiên xuất hiện trong đầu chúng ta có lẽ là giảm tử vong do mệt mỏi của người lái xe.

Với một hệ thống có thể cung cấp cảnh báo chệch làn đường, có thể tránh được những tai nạn lớn và cứu được tính mạng của người lái xe, đồng hành và người đi đường.

Tính chính xác của hệ thống nằm ở chỗ đào tạo thuật toán sử dụng một bộ sưu tập các hình ảnh. Tuy nhiên, việc phát triển một DDS mạnh mẽ là không thể nếu các khung mắt không được chụp đúng cách và hệ thống không được đào tạo trên các tập dữ liệu lớn. Hơn nữa, việc xác định vị trí trên mắt có thể trở nên khó khăn nếu người lái xe đeo các chướng ngại vật như kính bảo hộ hoặc mũ. [9]

V. CÁCH THỨC HOẠT ĐỘNG CỦA MÔ HÌNH PHÁT HIỆN BUỒN NGỦ

Thực tế cho thấy, việc phát hiện và phân loại đối tượng nói chung và việc nhận dạng buồn ngủ nói riêng trực tiếp từ hệ thống camera giám sát hoặc webcam hiện vẫn là một bài toán phức tạp, còn nhiều khó khăn, thách thức.

Webcam/Camera thu nhận hình ảnh

Phân đoạn video

Frame

Phát hiện mặt người

Khuôn mặt

Tiền xử lý,

Trích chọn đối tượng

Phân loại buồn ngủ, không buồn ngủ

Hình 2

***Hình 5*.** *Phương pháp phát hiện buồn ngủ từ webcam/camera*.

Một trong những thách thức điển hình của bài toán này là hình ảnh khuôn mặt của đối tượng chuyển động và thay đổi liên tục với nhiều tư thế góc nghiêng / xoay và trạng thái khác nhau. Điều này đòi hỏi các thuật toán nhận dạng phải có khả năng tổng quát hoá, không bị ảnh hưởng nhiều bởi độ nghiêng / xoay, và dịch chuyển của đối tượng. Ngoài ra, việc nhận dạng trực tiếp từ camera/webcam luôn đòi hỏi phải đáp ứng được tính thời gian thực (real time) từ webcam/camera được đề xuất cụ thể trên *Hình 5*.

Từ tín hiệu video đầu vào, bước xử lý đầu tiên sẽ tiến hành phân đoạn video thành các khung hình (frame) riêng biệt. Việc phân đoạn video ở đây được tiến hành theo thời gian (ngưỡng được chọn hiện tại là 24 khung hình trên giây). Mỗi khung hình có thể không hứa, chứa một phần hoặc chứa toàn bộ khuôn mặt. Vì vậy, trong bước xử lý đầu tiên, thuật toán sẽ tiến hành phát hiện (face detection) và xác định vị trí các đối tượng (trong trường hợp này là vùng Mắt) trên ảnh. Các đối tượng được phát hiện sau đó sẽ tiếp tục được tiền xử lý nhằm tăng cường chất lượng hình ảnh (loại nhiễu, khử bóng mờ), chuẩn hoá kích thước và độ phân giải ảnh, căn chỉnh đối tượng về hướng trực diện. Các khuôn mặt sau khi đã tiền xử lý sẽ được sử dụng làm đầu vào cho một mô hình mạng Neural học sâu (DNN-Deep Neural Network). Mô hình này sẽ tự động học và trích lọc ra những đặc trưng để phân lớp đối tượng. Bước xử lý cuối của thuật toán sẽ tiến hành phân lớp các đối tượng mắt nhắm/mở. Để thực hiện được điều này, các mô hình phân lớp cần phải được huấn luyện với một tập mẫu cho trước. Trong đó, mỗi mẫu khuôn mặt được thể hiện đặc trưng DNN ở bước trên.

A collage of a person's face

Description automatically generated

***Hình 12****. Tập ảnh dùng để huấn luyện máy phát hiện mắt nhắm.*

A collage of a person's face

Description automatically generated with medium confidence

***Hình 13****. Tập ảnh dùng để huấn luyện máy phát hiện mắt mở.*

Dự án sử dụng Drowsiness\_dataset có mặt trên nền tảng Kaggle. Bộ dữ liệu ban đầu chứa bốn lớp để phân loại hình ảnh thành Open Eyes, Closed Eyes, Yawning hoặc No-Yawning. Tuy nhiên, phạm vi của nghiên cứu này là phân loại buồn ngủ dựa trên việc mắt nhắm hay mở. Vì vậy, tôi sẽ chỉ sử dụng hai lớp của bộ dữ liệu. Đặc điểm của bộ dữ liệu như sau:

* Bộ dữ liệu chứa tổng cộng 1452 hình ảnh trong loại mắt mở (*Hình 12*).
* Bộ dữ liệu chứa tổng cộng 726 hình ảnh trong loại mắt nhắm (*Hình 13*).
* Bộ dữ liệu đã được cân bằng, vì cậy không cần phải làm cân bằng bộ dữ liệu.
* Nhãn lớp sẽ được học: ‘Open’ và ‘Closed’.
* Nhãn lớp được mã hoá sao cho 0 đại diện cho Mắt mở và 1 minh hoạ Mắt nhắm.

Hình ảnh được đưa qua các bước xử lý để làm cho kích thước của mỗi hình ảnh bằng nhau (32, 32, 3). Sau đó, bộ dữ liệu được chia thành Train set và Test set theo tỷ lệ 80-20.

Nhiệm vụ chính của chương trình là phát hiện tình trạng buồn ngủ trong quá trình lái xe. Ngoài ra, chường trình còn thực hiện công việc phát hiện, tách các khuôn mặt người (nếu có) từ một frame ảnh thu được từ camera.

Như vậy, các chức năng của chương trình bao gồm:

* Dựa trên dữ liệu đa được train là các tệp ảnh được lưu trong dataset như Hình 6, Hình 7. Chương trình sẽ kết nối đến webcam, hiển thị lên ImageBox.
* Thực hiện tách các khuôn mặt trên ImageBox.
* So sánh các đặc điểm vừa tách được với các đặc điểm đã được train.
* Kết quả hiển thị “OPEN” nếu phát hiện mắt mở hoặc “CLOSED” nếu phát hiện mắt nhắm.

Đầu vào của hệ thống là 1 ảnh chứa khuôn mặt cần xử lý, ảnh này có thể là ảnh tĩnh, lấy từ trong bộ nhớ máy tính hoặc frame ảnh bắt được từ dòng hình ảnh của camera. Sau khi có được ảnh đầu vào thì bắt đầu tiến hành tìm kiếm, phát hiện các khuôn mặt trong ảnh có đeo khẩu trang hay không.

Diagram

Description automatically generated

***Hình 14****. Lưu đồ giải thuật chọn ảnh đầu vào*

Sau khi nhận được ảnh đầu vào, hệ thống sẽ thực hiện chức năng phát hiện buồn ngủ trong ảnh. Như đã nói trong phần giới thiệu, bài toán phát hiện buồn ngủ là một bài toán khó, nên ta không đi sâu tìm hiểu cách giải quyết bài toán này.

A picture containing diagram

Description automatically generated

***Hình 15****. Cách thức phát hiện mắt nhắm/mở.*

Ở đây ta sử dụng phương pháp tiếp cận theo hướng diện mạo (appearance-based) được thực hiện nhanh bằng thuật toán adaboost thông qua hàm cvHaarObjects của bộ thư viện OpenCV. Hàm này thực hiện việc phát triển đối tượng dựa trên các đặc trưng haar-like, cụ thể là nhờ vào một bộ Cascade được truyền vào cho hàm. Bộ cascade được xây dựng theo dạng tree-node và đã được huấn luyện từ trước.

Việc huấn luyện bộ Cascade có thể thực hiện từ những dữ liệu thu thập được để phục vụ cho quá trình nhận dạng. Ví dụ, muốn phát hiện một người có buồn ngủ hay không, ta thu thập các ảnh vùng mắt ở nhiều góc độ, điều kiện ánh sáng khác nhau, sau đó cho bộ phận dạng học theo thuật toán của CNN training. Kết quả thu được sẽ có một mô hình nhận diện, được sử dụng để nhận dạng các đối tượng.

Việc đào tạo mô hình chạy tổng cộng 20 EPOCH với quy mô (BATCH SIZE) là 128. Lớp ImageDataGenerator được sử dụng để ngẫu nhiên hoá các hình ảnh đào tạo để có hiệu suất tót hơn của mô hình.

Trong quá trình nghiên cứu, chúng tôi đã chọn hai mô hình để huấn luyện đó là mô hình CNN và mô hình MobilNetV2.

1. *Huấn luyện trên kiến trúc CNN*

Với mô hình CNN, kiến trúc được xây dựng như sau:

* Convolution2D layer: 32 nodes, kernel\_size = 3.
* MaxPooling2D: pool\_size = 1.
* Convolution2D layer: 32 nodes, kenel\_size = 3.
* MaxPooling2D: pool\_size = 1.
* Convolution2D layer: 64 nodes, kenel\_size = 3.
* MaxPooling2D: pool\_size = 1.
* Fully connected layer: 128 nodes.
* Flatten layer.
* Output Fully connected layer (Dense):2 nodes.

Hình ảnh đầu vào chứa 1024 điểm ảnh (24 x 24). Lớp phức hợp được hạ thấp bởi Pooling layer. Mỗi lớp nhận được một khối lượng đầu vào 3D, được gọi là feature-map và chuyển nó sang dạng khác. Bằng cách xếp chồng các lớp và mẫu xuống các đầu ra. Các lớp cuối cùng của CNN là các lớp được kết nối đầy đủ theo kiểu chuẩn. Các lớp này kết hợp các bộ mô tả cuối cùng của các hình ảnh đầu vào, có thể được coi là ccá đại diện toàn cục của hình ảnh, hoặc chúng phân loại các hình ảnh đầu vào thành các lớp tuỳ thuộc vào một hàm mục tiêu.

Chart, scatter chart

Description automatically generated

***Hình 16****. Các điểm ảnh trên khuôn mặt.*

Các convolutional layer có các parameter (kenel) đã được học để tự điều chỉnh lấy ra những thông tin chính xác nhất mà không cần chọn các đặc trưng.

MaxPooling2D thường được sử dụng ngay sau lớp Copnvolution2D để đơn giản hoá thông tin đầu ra để giảm bớt số lượng neural. Như vậy, qua lớp Max Pooling thì số lượng neural giảm đi phân nửa. Trong một mạng CNN có nhiều Feature map, chúng ta sẽ cho mỗi MaxPooling khác nhau. Chúng ta có thể thấy rằng Max Pooling là cáh hỏi xem trong các đặc trưng này thì đặc trưng nào là đặc trưng nhất.

Flatten là một hành động reshape cụ thể trong đó tất cả các axes được làm phẳng hoặc ghép lại với nhau. Nói dễ hiểu là nó loại bỏ tất cả các kích thước ngoại trừ một kích thước.

Dense layer hay Fully connected layer là một lớp cổ điển trong mạng neural nhân tạo. Mỗi neural nhận đầu vào từ tất cả neural lớp trước đó.

Xác định mô hình bằng keras:

Graphical user interface, application, Word

Description automatically generated

***Hình 17****. Mô hình được xác định bằng keras.*

Chart, line chart

Description automatically generated

***Hình 18****. Đánh giá validation accuracy kiến trúc CNN.*

Sau quá trình huấn luyện và chỉnh sửa tham số, cho ra kết quả cuối cùng với độ chính xác 97,24%.

*B. Huấn luyện trên kiến trúc MobileNetV2*

Text

Description automatically generated

***Hình 19****. Kiến trúc MobileNetV2.*

Ta thấy xuất hiện MobilêNtV2. Kiến trúc MobileNetV2 dựa trên cấu trúc phần dư ngược, trong đó đầu vào và đầu ra của khối dư là các lớp nút cổ chai móng đối diện với các mô hình phần dư truyền thống sử dụng các biểu diễn mở rộng trong đầu vào. MobileNetV2 sử dụng phức hợp theo chiều sâu nhẹ để lọc các tính năng trong lớp mở rộng trung gian. Ngoài ra, các đặc điểm không tuyến tính trong các lớp hẹp đã bị loại bỏ để duy trì sức mạnh đại diện.

Graphical user interface, application, website

Description automatically generated

***Hình 20****. Cách thức hoạt động của MobileNetV2.*

AveragePooling2D là lớp hoạt động tổng hợp trung bình cho dữ liệu không gian. Đối số pool\_size là một số nguyên hoặc tuple của 2 số nguyên, các yếu tố để giảm quy mô (dọc, ngang). Sẽ giảm một nửa đầu vào trong cả hai kích thước không gian. Nếu chỉ có một số ngjuyên được chỉ định, cùng một độ dài cửa sổ sẽ được sử dụng cho cả hai kích thước.

Chart

Description automatically generated

***Hình 21****. Đánh giá validation accurracy kiến trúc MobileNetV2.*

Kết quả thực nghiệm (Hình ) cho thấy, việc sử dụng các lớp mạng tích chập (CNN) để trích chọn đặc trưng cho độ chính xác cao. Do thuật toán được huấn luyện với tập dữ liệu lớn, đa dạng nên các đặc trưng phát hiện được thường ít bị ảnh hưởng bởi nhiễu và các tính chất nghiêng, xoay của ảnh. Tuy nhiên, do mạng được kiến trúc nhiều lớp và số liên kết giữa các lớp mạng rất lớn nên việc tính toán trên mạng thường mất nhiều thời gian. Điều này dẫn tới tốc độ tổng thể của thuật toán nhận dạng sẽ bị ảnh hưởng. Vì vậy, để đảm bảo thuật toán có thể đáp ứng tính thời gian thực (real-time) trong quá trình nhận dạng, chúng tôi đã tận dụng khả năng tính toán GPU, cho phép việc tính toán trên các lớp mạng thực hiện theo cơ chế song song.

V. GIỚI THIỆU TRIỂN KHAI VÀ ĐÁNH GIÁ GIẢI PHÁP

*A.Giới thiệu triển khai.*

Sau quá trình huấn luyện (training), chúng tôi tiến hành xây dựng một ứng dụng mô phỏng camera giám sát trên xe ô tô bằng webcam và ứng dụng desktop.

A picture containing diagram

Description automatically generated

***Hình 22****. Camera giám sát buồn ngủ được lắp đặt trong xe.*

*B.Đánh giá thực nghiệm.*

Chương trình thực nghiệm được cài đặt trong môi trường python, sử dụng các thư viện Keras cho việc biểu diễn, lưu trữ và thao tác dữ liệu, thư viện OpenCV để thực hiện các thao tác xử lý ảnh cơ bản, thử nghiệm các mô hình học máy. Chường trình được thử nghiệm trên hệ điều hành Windows 11, máy PC tốc độ 2,4GHz, bộ nhớ 16GB.

Hiệu quả của mô hình nhận dạng được đánh giá trên các bộ cơ sở dữ liệu chuẩn (chứa các khung hình được thu thập thừ các thiết bị camera, webcam khác nhau), được công bố dùng chung cho các nhóm nghiên cứu trên thế giới. Gồm 2 tập dữ liệu: Open, Closed. Ảnh trong cơ sở dữ liệu là ảnh màu 24 bit. Định dạng PNG, JPEG, Tập dữ liệu chứa một tập hợp các hình ảnh khuôn mặt chủ yếu được thực hiện bởi các sinh viên đại học năm đầu tiên có độ tuổi từ 18 đến 20 và một số người lớn tuổi, một số cá nhân đeo kính, mắt thâm, mắt hí, nhiều chủng tộc khác nhau,…

Quá trình đánh giá thực nghiệm được chia thành 2 công đoạn: Đánh giá hiệu quả của mô hình phát hiện buồn ngủ được đánh giá dựa trên các độ đo được định nghĩa cụ thể trong phần sau đây:

* Độ chính xác phát hiện khuôn mặt DP (Detection Precision):

DP = Số vùng khuôn mặt phát hiện đúng trên tổng số khuôn mặt cần phát hiện.

* Khả năng tìm hết DR (Dêtction Recall):

DR = Số vùng khuôn mặt phát hiện đúng / (Số vùng khuôn mặt phát hiện đúng + số vùng khuôn mặt không được phát hiện).

* Độ trung bình điều hoà DM (Detection F-Measure):

DM = (2Xfdp\*FDR)/(FDP+FDR).

Bên cạnh đó, để các kết quả thực nghiệm chính xác và trực quan, trong quá trình thử nghiệm, chúng tôi đã so sánh hiệu quả của mô hình phát hiện khuôn mặt đề xuất với mô hình phát hiện khuôn mặt sử dụng đặc trưng Haar-Like và bộ phân lớp AdaBoost (Được quy ước gọi tên là phương pháp Haar-Like AdaBoost). Các kết quả thực nghiệm được mô tả cụ thể trên Bảng 1.

*Bảng 1. Đánh giá hiệu quả phát hiện buồn ngủ.*

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dữ liệu thử nghiệm | Số mẫu | Phương pháp đề xuất | | | Phương pháp Haar-Like AdaBoost | | |
| DP | DR | DM | DP | DR | DM |
| Open | 650 | 98.43 | 98.43 | 98.43 | 80.71 | 80.71 | 80.71 |
| Close | 700 | 97.43 | 97.43 | 97.43 | 79.43 | 79.43 | 79.43 |

Hiệu quả của mô hình nhận dạng tổng thể được đánh giá dựa trên độ chính xác nhận dạng, được định nghĩa cụ thể như sau:

R\_Precision = Số khuôn mặt nhận dạng đúng / Tổng số khuôn mặt cần nhận dạng.

Quá trình đánh giá thực nghiệm được thực hiện lần lượt trên từng tập dữ liệu. Mỗi tập dữ liệu được chia ngẫu nhiên thành 2 tập training và testing theo tỷ lên 80/20.

Việc huấn luyện mô hình gôm 2 công đoạn: Huấn luyện bộ trích chọn đặc trưng (mô hình mạng neural tích chập) và huấn luyện bộ phân lớp SVM. Quy trình huấn luyện được tiến hành cụ thể như sau: Từ tập mẫu huấn luyện đầu vào, trước tiên bộ phát hiện khuôn mặt sẽ tiến hành tìm kiếm, định vị và crop vùng ảnh khuôn mặt trên mỗi khung hình. Toàn bộ tập ảnh khuôn mặt crop sau đó sẽ được sử dụng làm đầu vào để huấn luyện mô hình trích chọn đặc trưng. Tập đặc trưng đầu ra của mô hình trích chọn đặc trưng sẽ được sử dụng làm đầu vào để huấn luyện mô hình phân lớp SVM.

Các kết quả thực nghiệm được mô tả cụ thể trên Bảng 2. Trong đó, hiệu quả của mô hình đề xuất được so sánh với phương pháp phân lớp sử dụng đặc trưng PCA và bộ phân lớp Eigenface (được quy ước gọi tên là phương pháp PCA-Eigenface).

*Bảng 2. Đánh giá độ chính xác nhận dạng.*

|  |  |  |  |
| --- | --- | --- | --- |
| Tập dữ liệu thử nghiệm | Số khuôn mặt cần nhận dạng | R\_Precision (%) | |
| PP đề xuất | PCA-Eigenface |
| Open | 650 | 98.6 | 82.12 |
| Close | 700 | 97.55 | 75.43 |

Từ các kết quả thực nghiệm cho thấy phương pháp đề xuất đạt được độ chính xác cao (Trên 97%) trên tất cả các tập dữ liệu thử nghiệm. Trong khi đó độ chính xác của phương pháp PCA – Eigenface bị ảnh hưởng nhiều ởi độ sáng và độ địch chuyển của ảnh đầu vào.

VII. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI

Bài báo này đề xuất một mô hình tổng thể cho việc phát hiện buồn ngủ từ webcam/camera. Trong đó tập trung chính vào công đoạn phát hiện và nhận dạng khuôn mặt. Hiệu quả của mô hình đã được đánh giá trên các tập dữ liệu chuẩn, dùng chung cho cộng đồng nghiên cứu về nhận dạng trên thế giới. Quá trình đánh giá thực nghiệm được chia thành 2 bước, trong đó hiệu quả của phương pháp phát triển khuôn mặt được đánh giá dựa trên 3 độ đo là độ chính xác (Precision), khả năng tìm hết (Recall) và độ đo (F-meansure), hiệu quả của mô hình nhận dạng khuôn mặt được đánh giá dựa trên độ chính xác nhận dạng. Các kết quả thực nghiệm cho thấy mô hình đề xuất đạt được độ chính xác cao và ổn định trong môi trường thực tế, có thể ứng dụng mô hình để giải quyết các bài toán ứng dụng điển hình như trên hệ thống camera giám sát,…

VI. TÀI LIỆU THAM KHẢO

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  |  | | --- | --- | | [1] | NHTSA, "Drowsy Driving," [Online]. Available: https://www.nhtsa.gov/risky-driving/drowsy-driving. | | [2] | Czeisler, "Drowsy Driving," in *Sleep Health 2016*, pp. 94-99. | | [3] | "VINMEC," [Online]. Available: https://www.vinmec.com/vi/tin-tuc/thong-tin-suc-khoe/suc-khoe-tong-quat/vi-sao-chung-ta-co-cam-giac-buon-ngu/. | | [4] | "Driver Fatigue and Road Accidents," in *A Literature Review and Position Paper*. | | [5] | J. H. a. R. Picard, "SmartCar: Detecting driver stress," in *Pattern Recognition*, pp. 218-221. | | [6] | M. U. a. J. S. A. Kircher, "Vehicle control and drowsiness". | | [7] | Anon, "“Perclos and eyetracking: Challenge and opportunity," Applied Science Laboratories, [Online]. Available: http://www.a-s-l.com. | | [8] | C. F.Xiao, "Yawning detection based on gabor wavelets and LDA," pp. 409-413. | | [9] | "DDS," Shaip, [Online]. Available: https://vi.shaip.com/. | |