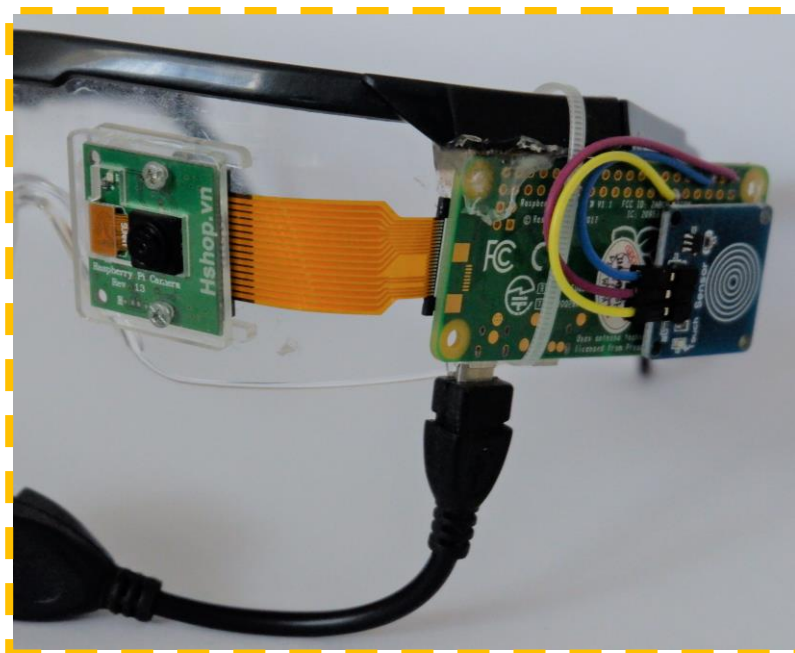


**CUỘC THI SÁNG TẠO KHOA HỌC, KỸ THUẬT
DÀNH CHO HỌC SINH TRUNG HỌC 2018 – 2019**

Mã số:

BÁO CÁO DỰ ÁN KHOA HỌC



**Dự án: THIẾT BỊ PHÂN TÍCH DỮ LIỆU TỪ
CAMERA DỰA VÀO THỊ GIÁC MÁY TÍNH
VÀ TRÍ TUỆ NHÂN TẠO**

Lĩnh vực : Hệ thống nhúng

MỤC LỤC

I.	Lý do lựa chọn đề tài	3
II.	Mục tiêu nghiên cứu và giới thiệu chung.....	3
1.	Mục tiêu nghiên cứu	3
2.	Giới thiệu chung.....	3
3.	Ý nghĩa	4
III.	Mô tả tổng quan.....	4
1.	Phần cứng.....	4
2.	Phần mềm.....	5
IV.	Các tính năng.....	7
1.	Miêu tả cảnh bằng câu văn có ý nghĩa.....	7
2.	Nhận diện vật thể.....	8
3.	Nhận diện gương mặt, miêu tả gương mặt.....	9
4.	Nhận diện ký tự.....	10
5.	Tra cứu các thông tin trên Internet	11
6.	Điều khiển thiết bị IoT.....	11
7.	Các tính năng mở rộng.....	12
8.	Giao tiếp giữa thiết bị và máy chủ	12
V.	Kết luận.....	13
1.	Ưu nhược điểm	13
2.	Đánh giá kết quả thử nghiệm với lý thuyết	13
3.	Kết luận	14
VI.	Tài liệu tham khảo	15

LỜI CẢM ƠN

Em xin gửi lời cảm ơn chân thành đến các thầy cô giáo trong Ban giám hiệu nhà trường, đặc biệt là thầy cô hướng dẫn đã truyền đạt những kiến thức, kinh nghiệm quý báu để em có thể hoàn thành được đề tài của mình một cách tốt nhất.

Em xin bày tỏ lòng biết ơn đến những tác giả của các bài báo cáo, bài viết cũng như các tác giả của những mã nguồn mở được em sử dụng trong đề tài này. Em xin cảm ơn những nhà cung cấp API, các dịch vụ đã cung cấp dữ liệu, tài nguyên để đề tài này được hoàn thành trọn vẹn.

Em xin cảm ơn gia đình, bạn bè đã luôn động viên, ủng hộ và khích lệ, tạo điều kiện giúp đỡ em trong quá trình thực hiện, hoàn thành sản phẩm và bài báo cáo này.

I. Lý do lựa chọn đề tài:

Trong thời kỳ bước vào cuộc cách mạng 4.0 với trí tuệ nhân tạo, dữ liệu lớn và vạn vật kết nối đang phát triển nhanh chóng. Xã hội ngày càng có nhiều vấn đề cần đến sự giúp ích của các thiết bị thông minh. Nhiều vấn đề xã hội nảy sinh như giám sát, phát hiện tội phạm, quản lý con người, hàng hóa, hỗ trợ người khuyết tật, ... Thị giác máy tính, trí tuệ nhân tạo, dữ liệu lớn, vạn vật kết nối sinh ra để giúp ích và giải quyết các vấn đề trên và có thể làm nhiều hơn thế nữa. Xuất phát từ mục đích đó, việc nghiên cứu và ứng dụng bằng cách kết hợp, xử lý, sử dụng hợp lý công nghệ trên tạo thành các thiết bị giúp ích cho con người là điều cần thiết.



Hiện nay, Việt Nam có gần 30 vạn người khiếm thị và gần 2 triệu người có vấn đề về thị giác. Người khiếm thị gặp rất nhiều cản trở trong cuộc sống. Họ cần có thiết bị thông minh để hỗ trợ cho họ chủ động hơn trong sinh hoạt và công việc hằng ngày. Tuy nhiên các thiết bị thông minh trên thị trường hiện nay như máy chuyển thể văn bản thành chữ nổi, kính thông minh, ... là các sản phẩm của nước ngoài, không hỗ trợ tiếng Việt và giá thành cực kỳ cao nên đa số người khiếm thị nước ta chưa tiếp cận được với các thiết bị này.

Xuất phát từ thực tế đó và từ việc nghiên cứu, tìm hiểu, tác giả mong muốn tạo ra một sản phẩm thực tế bằng những lý thuyết nghiên cứu được để giúp ích cho những người khiếm thị thiếu đi ánh sáng và cần một trợ lý thông minh để họ dễ dàng hòa nhập và gần gũi với các công nghệ của thời đại 4.0 hơn.

II. Mục tiêu nghiên cứu và giới thiệu chung:

1. Mục tiêu nghiên cứu:

- Kết hợp thị giác máy tính và trí tuệ nhân tạo nhằm tạo ra thiết bị có khả năng nhận diện khuôn mặt con người, vật thể, kí tự bằng quang học và có khả năng miêu tả khuôn mặt, cảnh vật.
- Tối ưu và sử dụng các kỹ thuật để có kết quả nhận diện chính xác hơn với các thiết bị sẵn có trên thị trường.
- Nghiên cứu nhận diện giọng nói, tổng hợp giọng nói, thiết kế hình dạng, hoạt động trơn tru, nhanh và điều khiển thiết bị dễ dàng.
- Ngoài những vấn đề trên, khi tạo dựng một thiết bị trợ giúp cho người khiếm thị cần nghiên cứu thêm việc sử dụng dữ liệu lớn để tìm kiếm, truy xuất các thông tin trên internet như tin tức, kiến thức và vạn vật kết nối để có thể điều khiển các thiết bị trong mạng lưới IoT.

2. Giới thiệu chung:

“Thiết bị phân tích dữ liệu từ camera dựa vào thị giác máy tính và trí tuệ nhân tạo” là thiết bị ứng dụng những tìm hiểu và nghiên cứu về thị giác máy tính, trí tuệ nhân tạo, dữ liệu lớn và vạn vật kết nối (IoT) của tác giả bao gồm

các tính năng: Nhận diện khuôn mặt, miêu tả gương mặt, nhận diện vật thể, miêu tả cảnh vật bằng một câu có nghĩa, nhận diện ký tự bằng quang học, tìm kiếm các thông tin, tin tức và điều khiển các thiết bị thông minh.

Mỗi tính năng mà tác giả nghiên cứu có thể ứng dụng để tạo ra các thiết bị khác nhằm giải quyết các vấn đề xã hội hiện nay. Và, với những nghiên cứu và tìm hiểu đó, tác giả đã tạo ra thiết bị là sản phẩm dành cho người khiếm thị, giúp họ có thể hình dung được con người, cảnh vật xung quanh, tìm kiếm, tiếp cận thông tin từ internet và giúp họ dễ dàng hòa nhập vào cuộc sống.

Thiết bị bao gồm: một camera để chụp hình ảnh, một nút bấm để người dùng ra lệnh, một microphone để nhận diện lệnh bằng giọng nói, một loa để phát ra kết quả xử lý và một bo mạch nhúng để xử lý dữ liệu từ camera, nhận diện lệnh, gửi dữ liệu về máy chủ, nhận dữ liệu từ máy chủ và phát ra kết quả đã xử lý. Thiết bị có thể gắn lên kính hoặc cài áo. Người dùng dễ dàng mang theo bên người. Người dùng ra lệnh bằng giọng nói và nhận kết quả bằng giọng nói.

3. Ý nghĩa:

Ý nghĩa của những nghiên cứu về phân tích dữ liệu từ camera dùng thị giác máy tính và trí tuệ nhân tạo: Ứng dụng được trí thông minh nhân tạo và thị giác máy tính để xây dựng và tạo ra các thiết bị phân tích hình ảnh giúp ích cho cuộc sống, những hoàn cảnh khó khăn. Có thể ứng dụng từng nghiên cứu để tạo ra nhiều thiết bị với chức năng khác nhau để giải quyết các vấn đề trong xã hội như cảnh báo đường kẹt xe, quản lý tội phạm, điểm danh, chấm công, làm máy đọc chữ, ...

Ý nghĩa của thiết bị: Thiết bị là ứng dụng của thị giác máy tính, trí tuệ nhân tạo - các công nghệ trong thời đại 4.0 - để phục vụ những người khiếm thị đang khó khăn để hòa nhập, là một sản phẩm thuần Việt, đa tính năng với các tính năng gần gũi với nhu cầu người khiếm thị, dễ sử dụng, một người bạn đồng hành giúp đỡ cho người khiếm thị nói riêng và mọi người nói chung.

III. Mô tả tổng quan:

1. Phần cứng:

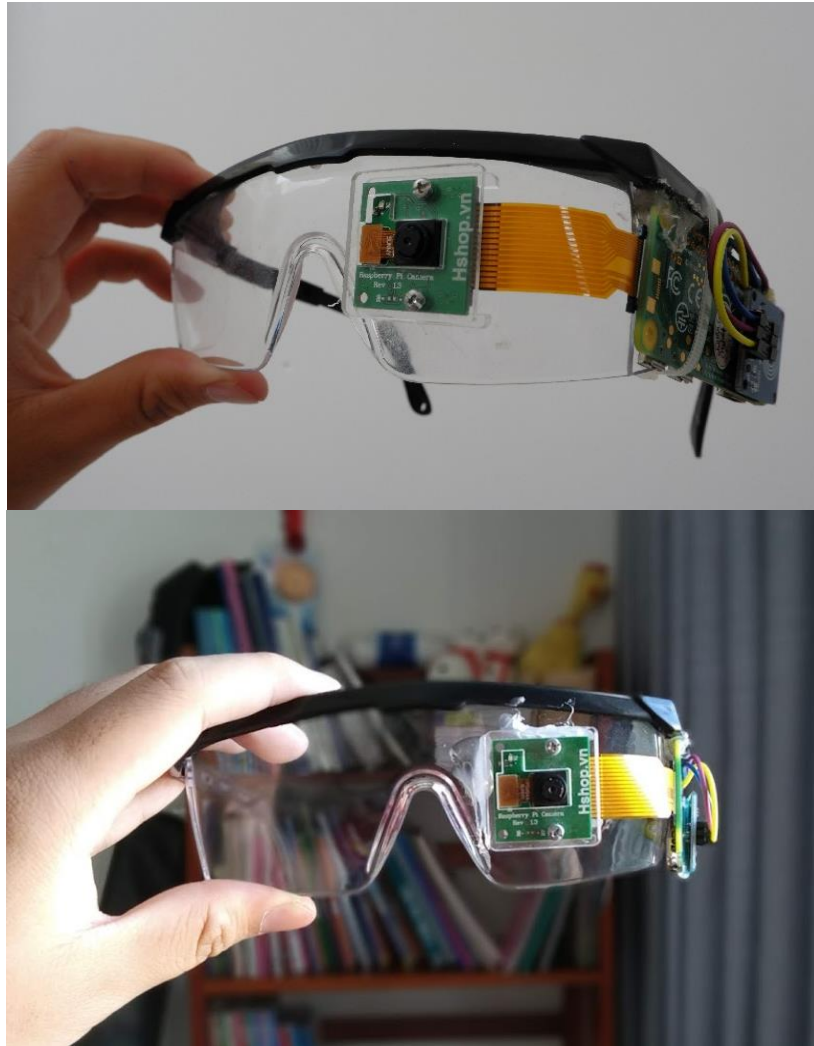
- Bo mạch nhúng Raspberry pi zero: máy tính nhúng nhỏ gọn, giá thành thấp, tính ổn định cao và cộng đồng hỗ trợ lớn.

- CPU: Broadcom BCM2835, tốc độ tối đa 1GHz.
- RAM: 512MB
- Nguồn: 5V, cung cấp qua cổng Micro USB
- Kích thước: 65mm x 30mm x 5mm
- Video & Audio: Hỗ trợ audio và video 1080P HD qua cổng mini

HDMI

- Raspberry Pi camera: tích hợp camera 5 Megapixel có độ nhạy sáng cao, có thể chụp tốt ở nhiều điều kiện ánh sáng khác nhau.

- Nút bấm cảm ứng: Module cảm biến chạm nhỏ gọn , dễ dàng sử dụng với 3 chân kết nối, nguồn điện cung cấp 2 ~ 5.5V.
- Microphone
- Loa



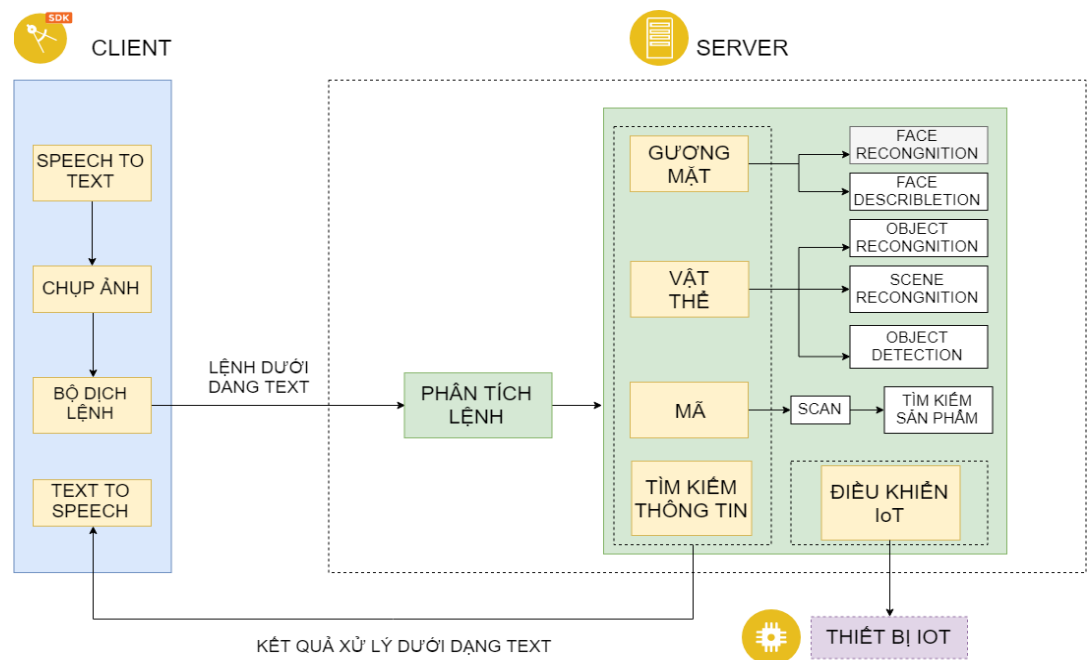
Hình ảnh của thiết bị

- Về mặt quang học:
 - + Camera đặt ở vị trí mắt của người dùng.
 - + Khoảng cách từ camera đến khuôn mặt, trang giấy cần nhận diện ký tự là 40-50cm. Đây là khoảng cách tối ưu vì vừa nằm trong dải tiêu cự của camera Raspberry Pi, vừa đảm bảo DPI của hình. Còn đối với vật thể thì vật thể cần nằm trọn trong khung hình.

2. Phần mềm:

- Sử dụng ngôn ngữ lập trình Python, C++, thiết bị và máy chủ đều sử dụng hệ điều hành Linux.
- Tính thân thiện với người dùng:
 - + Bấm nút để bắt đầu ra lệnh cho thiết bị
 - + Điều khiển thiết bị bằng giọng nói và thu được kết quả là giọng nói.

- + Các bước xử lý dữ liệu được gửi về máy chủ, thiết bị chỉ chụp hình ảnh, nhận diện lệnh, gửi về máy chủ và nhận kết quả từ máy chủ. Thiết bị thực hiện tác vụ nhẹ nên ít bị nóng.
- Cấu trúc và quy trình hoạt động:



Thiết bị (với nhiệm vụ là Client) sau khi nhấn nút và thu nhận lệnh thì sẽ chụp lại hình ảnh, gửi lệnh và hình ảnh đến máy chủ (server) để xử lý các tác vụ nhận diện, tìm kiếm thông tin cũng, tiền xử lý, gửi lên các API, hậu xử lý. Máy chủ sẽ gửi kết quả xử lý về cho thiết bị, thiết bị sẽ phát ra giọng nói cho người dùng hiểu.

- Các mã nguồn mở, framework được sử dụng:

- + Thị giác máy tính được dùng để thu nhận hình ảnh, thực hiện các bước tiền xử lý, mã hóa và giải mã hình ảnh. Sử dụng thư viện OpenCV – là một thư viện mạnh về xử lý hình ảnh, cộng đồng sử dụng lớn mạnh và có khả năng sử dụng các models Caffè, TF,...

- + Face_recognition là thư viện python được xây dựng với Dlib giúp phát hiện và nhận diện gương mặt với học sâu, một thư viện dễ dàng sử dụng và có hiệu quả cao.

- + Google Vision, Microsoft Vison là những api mạnh, cung cấp các dịch vụ xử lý hình ảnh. Lợi ích của việc sử dụng là không cần sở hữu máy chủ mạnh để huấn luyện, dataset được làm bởi những đội ngũ chuyên nghiệp, độ chính xác cao.

- + Keras là một framework hoạt động trên nền của TensorFlow, dễ sử dụng bởi cung cấp các lệnh sẵn, nhanh chóng tạo được một mạng Nơ ron từ dữ liệu có sẵn. Dữ liệu được tác giả khai thác để cho máy học từ ImageNet.

+ AutoML là công cụ xây dựng tự động mô hình deep learning trong việc nhận diện hình ảnh, sử dụng thuật toán tiên tiến để tự động tạo ra mô hình hiệu quả nhất.

+ Các thông tin về thời tiết, bài nhạc, bài báo được lấy từ các nguồn dữ liệu tin cậy: Weather.com, NCT.vn, VNexpress.net, ZingNews.vn,... Các định nghĩa được khai thác từ Google Knowledge Graph API, Wikipedia Api.

+ Text to speech và Speech to text là hai phần quan trọng đóng vai trò như phương thức giao tiếp giữa chương trình và người dùng. Sau quá trình chọn lọc thì Google cung cấp trình tạo giọng nói và nhận diện giọng nói tốt với độ ổn định và người dùng không phải xử lý nhiều về dấu câu. Đoạn văn được thu vào và tách thành từng câu và cho xử lý, tốc độ phản hồi sẽ nhanh hơn.

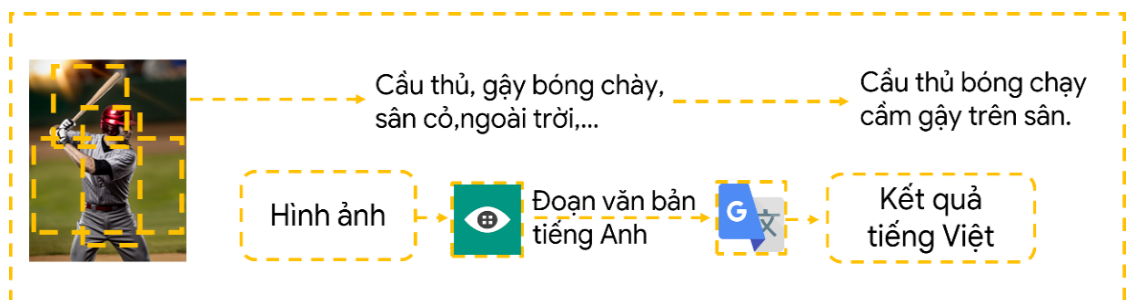
IV. Các tính năng:

1. Miêu tả cảnh bằng câu văn có ý nghĩa:

- Computer Vision API của Microsoft là một dịch vụ công nghệ phục vụ xử lý hình ảnh. Computer Vision API có khả năng miêu tả hình ảnh và tạo ra một câu đủ ngữ pháp và nghĩa để miêu tả bằng AI.

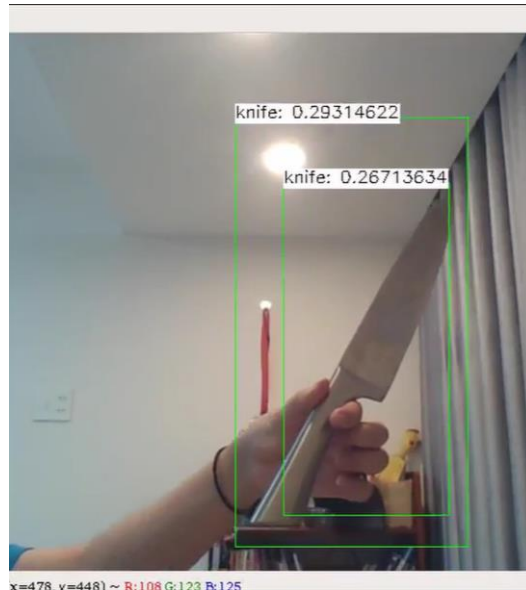
- Kết quả Computer Vision API cho ra là tiếng Anh. Để dịch kết quả từ tiếng Anh sang tiếng Việt sử dụng Google Translate API.

- Kết quả từ Computer Vision API có các câu miêu tả rất đặc trưng nhưng Google Translate có thể dịch sai nên cần được xử lý riêng như “a close of” (cận cảnh của), “piece of” (miếng), “on plate” (trên đĩa),...



2. Nhận diện vật thể:

- Sử dụng phương pháp phân loại vật thể bằng Deep learning. Bài toán đặt ra là nhận diện vật thể với tốc độ nhanh, sử dụng dữ liệu đầu vào là ảnh chụp, xử lý với mạng nơ-ron và đầu ra là nhãn của vật thể.



- Sử dụng mô hình được huấn luyện sẵn là MobileNet để nhận diện vật thể. Lợi ích khi dùng mô hình được huấn luyện sẵn là đã được huấn luyện trên tập dữ liệu ImageNet bởi các chuyên gia của Google, MobileNet được sinh ra cho các thiết bị di động nên có tốc độ xử lý nhanh và cần ít tài nguyên. Sử dụng OpenCV kết hợp với mô hình được huấn luyện sẵn MobileNet với việc nhận diện tới 90 vật thể. Bao gồm các vật thể nguy hiểm đối với người khiếm thị như dao, kéo, nĩa, các loại phương tiện giao thông.

- Thiết bị có khả năng nhận diện các vật bao gồm: phương tiện giao thông, động vật, các dụng cụ sinh hoạt hằng ngày, các loại đồ ăn, hoa quả. Các nhãn liên quan đến phương tiện giao thông, dụng cụ làm bếp sắc nhọn và các loại đồ ăn được ưu tiên hơn các nhãn khác bởi tận dụng việc nhận diện vật thể nhanh này có thể cảnh báo được nguy hiểm và thông tin cần thiết cho người khiếm thị với đồ vật họ tiếp xúc.

- Tốc độ nhận diện dựa vào cấu hình của máy chủ. Hiện nay, tốc độ nhận diện một bức hình là xấp xỉ 0.8 giây, có thể nhận diện theo thời gian thực. Vì tối ưu phần cứng để thiết bị có thể sử dụng lâu dài mà không quá nóng thì mỗi khi ra lệnh thiết bị mới bắt đầu nhận diện vật thể, song thời gian nhận diện không bị giảm.

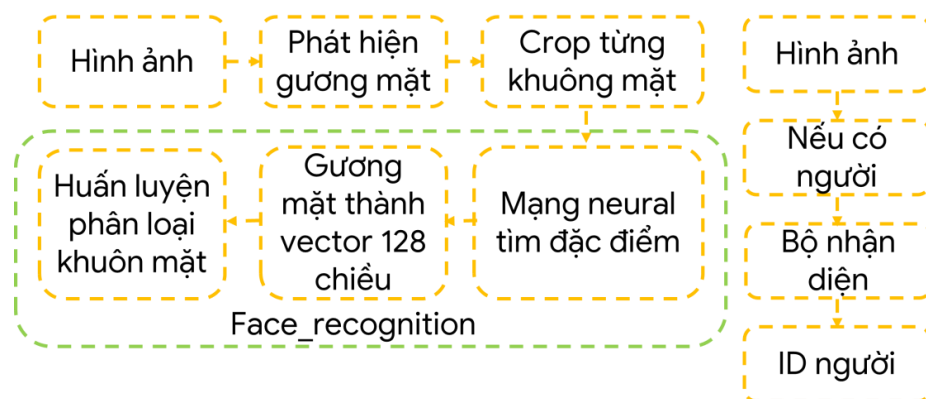
- Với việc nhận diện vật thể tùy chọn: Bài toán ở đây là nhận diện quần áo, tác giả sử dụng AutoML - với khả năng tự động xây dựng các mô hình Deep Learning cho bài toán Computer Vision (Thị giác máy tính). Với AutoML không cần am hiểu quá nhiều về Deep Learning nhưng vẫn có thể xây dựng được mô hình bởi sử dụng thuật toán Efficient Neural Architecture

Search khi mà máy tính thực hiện việc chọn các thành phần cần thiết từ một không gian xác định để xây dựng mạng nơron và sau đó cải thiện độ chính xác của nó.

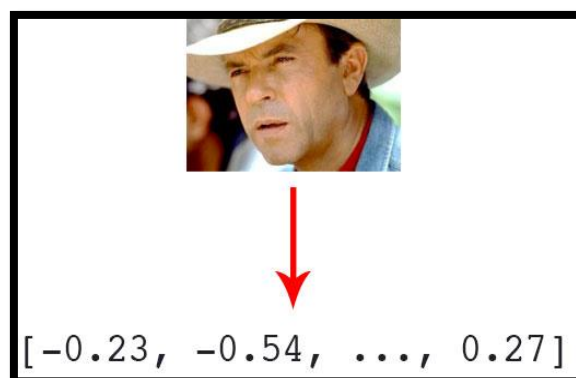


Các bước chuẩn bị dữ liệu, train và kiểm tra mô hình trên AutoML

3. Nhận diện gương mặt, miêu tả gương mặt:



- Nhận diện gương mặt sử dụng máy học có độ chính xác cao.

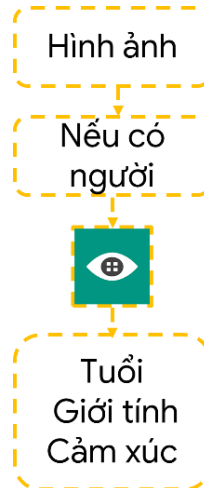


- Gương mặt được phân tích về các vector 128 chiều bằng deep learning. Sau đó máy sẽ lưu lại dữ liệu về khuôn mặt đó và thêm các thông tin cá nhân bao gồm: tên, số điện thoại (nếu được cung cấp để tạo thành danh bạ riêng trong máy). Mỗi gương mặt sẽ được đánh dấu bằng các chỉ số để tiện việc quản lý và truy xuất dữ liệu.

- Khi nhận diện thì gương mặt sẽ được so sánh với các gương mặt mà máy đã được học bằng Support Vector Machines (SVM) và đưa ra kết quả gương mặt gần giống nhất.

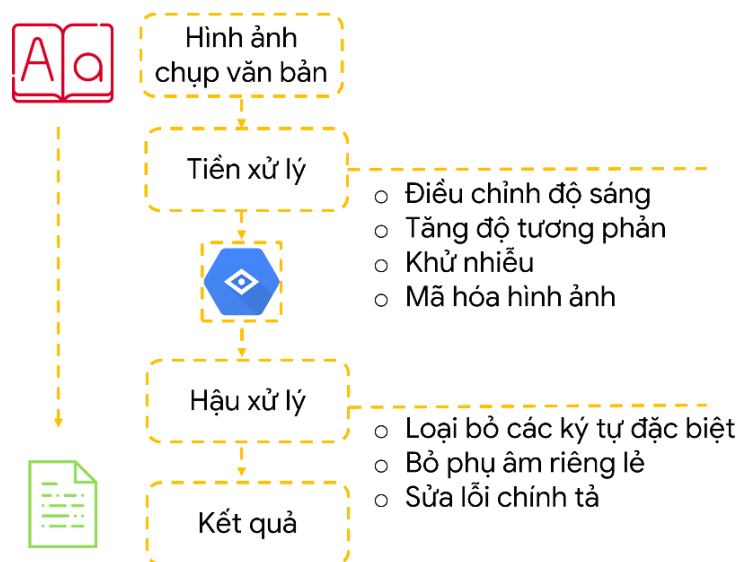
- Kết quả nhận diện chính xác cao và kể cả khi khuôn mặt thay đổi cảm xúc hay nghiêng nhẹ thì vẫn có thể nhận diện được.

- Miêu tả khuôn mặt sử dụng Face API của Microsoft. Kết quả phân tích một khuôn mặt cho ra bao gồm: tuổi, giới tính và cảm xúc như: vui, buồn, tức giận, bối rối, lo lắng.

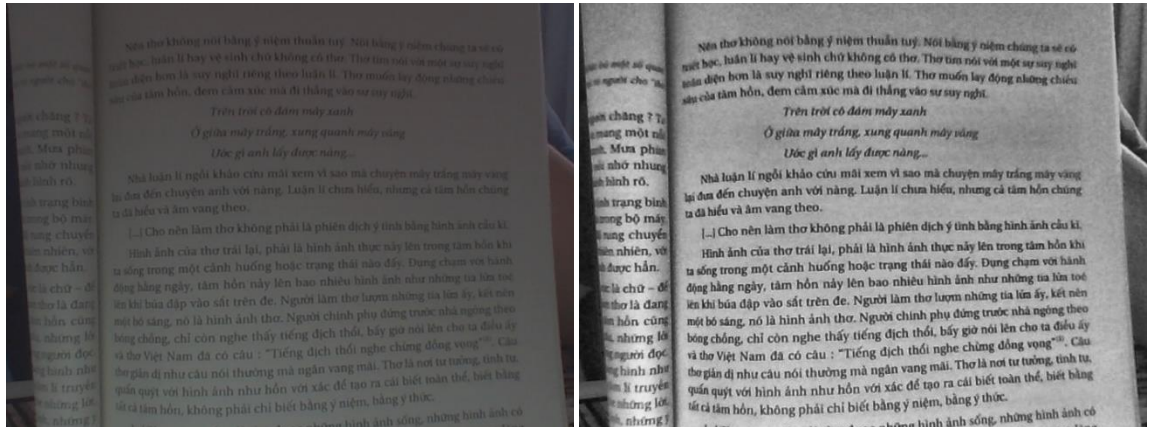


4. Nhận diện ký tự:

- Đọc kí tự bằng quang học (OCR) có nhiều công cụ để thực hiện nhưng Google Vision API là lựa chọn số một bởi độ hoàn thiện và tự động phân tách được các phần trong một trang văn bản, có khả năng dự đoán các từ tiếp theo và tự sửa lỗi chính tả. Việc sử dụng Google Vision cũng không cần yêu cầu phải huấn luyện lại bộ nhận dạng, dễ dàng sử dụng hơn rất nhiều so với việc sử dụng Tesseract. Tuy nhiên nhược điểm là yêu cầu thiết bị cần phải có kết nối mạng Internet.



- Tiền xử lý: Theo các trình tự xám hóa, khử nhiễu, điều chỉnh ánh sáng và tăng độ tương phản để việc xử lý nhận diện ký tự có độ chính xác cao hơn. Cân bằng sáng nhằm giảm chói hoặc thiếu sáng của bức ảnh, tuy nhiên chỉ trong khả năng có thể, nếu không bức hình sẽ bị nhiễu (noise). Thực hiện cân bằng sáng bằng Histogram Equalization để trải đều các pixel ở các vùng sáng/tối. Màu giấy và chữ thường có màu đối nghịch nên tăng độ tương phản để chữ nổi rõ ràng hơn. Dùng phân ngưỡng ảnh (threshold) để giảm các vùng tối, noise sau khi cân bằng sáng và tăng tương phản.



Hình ảnh sau khi qua bước tiền xử lý

- Hình ảnh được mã hóa thành chuỗi base64, gửi lên Google Vision API và nhận được kết quả là văn bản có trong hình đầu vào.

- Hậu xử lý: Sau khi nhận được kết quả OCR của Google Vision thì thiết bị sẽ tự bỏ các ký tự phụ âm đứng riêng lẻ, các dấu “,”,”’”,“:”,“?”,“!”,“(”,“),”,-”,“;” ra khỏi văn bản là kết quả sau khi nhận diện.

- Ngoài ra có thể dự đoán và sửa lỗi chính tả một số từ bằng cách tính toán xác suất xảy ra lỗi sai của từ đó với từ có thể sửa, tuy nhiên còn nhiều hạn chế. Vấn đề của Google Vision API khi thực hiện OCR không phải là nhằm thành chữ không tồn tại như “mg”, “gk”,... mà là thiếu dấu. Vậy nên ở đây tác giả sử dụng Vietnamese Accentizer của FPT để thêm dấu Tiếng Việt.

5. Tra cứu các thông tin trên Internet:

- Sử dụng Google knowledge graph, Wikipedia để khai thác Big Data về các khái niệm.

- Khi tìm kiếm định nghĩa thì hệ thống sẽ so sánh độ chính xác của yêu cầu với các vật thể tìm được. Xuất ra là kết quả chính xác nhất hoặc gần nhất về định nghĩa cần tìm.

6. Điều khiển thiết bị IoT:

- Người dùng có thể cài đặt và điều khiển các thiết bị đã kết nối trong mạng nội bộ với địa chỉ IP và tên thiết bị.

- Có thể điều khiển bóng đèn, quạt, thiết bị đo độ ẩm và nhiệt độ, công tắc điện,...

- Điều khiển qua giao thức http request.

7. Các tính năng mở rộng:

- Tra cứu thời tiết hôm nay: Sử dụng Api của OpenWeatherMap, thông tin lấy về bao gồm thời tiết, nhiệt độ, độ ẩm theo ngày hiện tại với vị trí mà người dùng yêu cầu.

- Đọc báo:

- + Sử dụng nguồn báo từ Google News, chia các bài báo thành từng nhóm 5 bài và người dùng có thể xem chọn lần lượt từng nhóm và từng bài bằng cách đọc thứ tự bài báo cần đọc. Ví dụ khi có lệnh đọc báo, thiết bị sẽ liệt kê ra tiêu đề bài báo từ thứ 1 đến thứ 5, người dùng có thể ra lệnh để thiết bị liệt kê tiếp tiêu đề bài báo thứ 6 đến thứ 10 và có thể yêu cầu thiết bị đọc bài với thứ tự được liệt kê.

- + Có thể tóm tắt được bài báo bằng cách sử dụng các thuật toán trong bài toán về xử lý ngôn ngữ tự nhiên. Các bước thực hiện bao gồm: tách câu trong văn bản, chuyển các câu sang dạng vector, phân cụm các câu trong văn bản và xây dựng tóm tắt. Hiện tại, có thể tóm tắt bài báo chỉ còn 4 đến 5 câu, tuy nhiên hạn chế là các bài viết phức tạp thì chưa đầy đủ ý.

8. Giao tiếp giữa thiết bị và máy chủ, nhận diện lệnh và đưa ra kết quả:

- Không xử lý trực tiếp trên bo mạch nhúng mà cần tới máy chủ vì bo mạch nhúng có kích thước nhỏ, không đủ mạnh cũng như sẽ nóng nếu thực hiện các tác vụ nặng, bởi vậy cần đến máy chủ để xử lý các tác vụ liên quan đến máy học thì kết quả sẽ nhanh hơn và giảm bớt sức nặng cho bo mạch nhúng.

- Giao tiếp giữa thiết bị và máy chủ: Khi người dùng nhấn nút ra lệnh, thiết bị sẽ tự động thu âm và chuyển từ giọng nói sang lệnh bằng Speech to Text của Google, đồng thời chụp lại hình ảnh qua camera, hình ảnh được mã hóa thành dạng base64. Thiết bị sẽ gửi về máy chủ thông qua giao thức http request. Máy chủ sẽ xử lý và gửi lại kết quả về thiết bị. Thiết bị sẽ tổng hợp giọng nói bằng Text to Speech của Google phát qua loa cho người dùng nhận.

- Các bước Text to Speech bao gồm gửi text lên API, tải đoạn âm thanh về và phát. Ở các yêu cầu nhận diện gương mặt và vật thể, do kết quả là cố định trong giới hạn dữ liệu mà thiết bị được cung cấp nên các file âm thanh đó sẽ được lưu lại. Như vậy, đến khi yêu cầu text to speech lại đoạn text đó, nếu có sẵn thì thiết bị sẽ phát lại đoạn âm thanh đã lưu sẵn, còn không sẽ tạo file âm thanh mới và lưu lại. Ví dụ như nhận diện một người tên “Khoa”, nếu lần đầu nhận diện thì file âm thanh “Khoa” sẽ được lưu lại và mỗi khi nhận diện một người tên “Khoa” thì thiết bị không cần phải chờ API Text to Speech xử lý, thời gian chờ sẽ được rút ngắn.

V. Kết luận:

1. Ưu nhược điểm:

- Ưu điểm:

- + Tự động ngắt thu nhận lệnh bằng âm thanh
- + Khả năng nhận dạng với độ chính xác tương đối cao
- + Hình ảnh được miêu tả bằng câu có nghĩa
- + Nhiều tính năng thông minh gần gũi với nhu cầu cuộc sống hằng ngày

- + Điều khiển được thiết bị IoT

- + Nhỏ gọn, linh hoạt có thể làm mắt kính hay đeo trước ngực.

- Nhược điểm:

- + Miêu tả cảnh mất tới gần 10 giây để xử lý
- + Cần có mạng internet để thiết bị hoạt động
- + Thiết bị còn chưa thẩm mỹ, chưa có case để bảo vệ thiết bị cũng như người dùng khỏi dây mạch.

2. Đánh giá kết quả thử nghiệm với lý thuyết:

- + Trọng lượng khoảng: 20g

- + Nguồn điện: 2A, 5V

- + Độ trễ: Thử với tốc độ mạng 17 Mps, thử 20 lần

	Độ trễ	Độ chính xác
Nhận diện giọng nói	~1.2 giây	90%
Tổng hợp tiếng nói đoạn văn 84 từ	~4 giây	N/A
Miêu tả cảnh	~7 giây	N/A
Nhận diện vật thể	~0.8 giây	80%
Nhận diện gương mặt	~0.5 giây	95%
Miêu tả gương mặt	~2 giây	65% (sai số lớn về dự đoán tuổi)
Đọc chữ OCR	~3.2 giây	85 %

3. Kết luận:

Hiện tại thiết bị chỉ ở mức thử nghiệm. Tuy nhiên, thiết bị có rất nhiều tính năng hữu ích cho người khiếm thị. Các nghiên cứu và việc ứng dụng các nghiên cứu vào thiết bị với mục tiêu giúp ích cho xã hội cho xã hội nói chung và người khiếm thị nói riêng bằng những công nghệ của thời đại 4.0, thuần Việt, cho người Việt. Thiết bị sẽ không ngừng phát triển, ngày càng chuyên sâu và tối ưu hơn, hoàn thành hơn để có thể giúp ích cho họ, giúp họ dễ dàng hòa nhập với cuộc sống thường ngày hơn.

Mã nguồn sẽ được mở cho cộng đồng cùng phát triển và xây dựng vì lợi ích chung của những người khiếm thị Việt Nam nói riêng và cộng đồng nói chung.

4. Hướng phát triển:

- Cải thiện độ chính xác trong việc nhận dạng
- Phát triển phần sửa lỗi sai khi nhận dạng ký tự bằng máy học
- Tối ưu hóa phần mềm với phần cứng
- Phát triển ứng dụng điện thoại để người dùng dễ dàng thêm tên mới vào danh bạ, cài đặt thiết bị dễ dàng
- Mở mã nguồn, xây dựng dữ liệu sách nói trực cho người khiếm thị
- Nghiên cứu thu gọn kính để trở nên thẩm mỹ hơn
- Tổng hợp giọng nói có âm điệu, dễ nghe.

VI. Tài liệu tham khảo:

- Tài liệu của Google Cloud Platform: <https://cloud.google.com/docs/>
- Tài liệu của OpenCV: <https://docs.opencv.org/>
- Tài liệu của Keras : <https://keras.io/documentation/>
- Tài liệu về TensorFlow: https://www.tensorflow.org/api_docs/
- Tài liệu về Underthesea: <https://github.com/undertheseanlp/underthesea/>
- Lưu trữ báo cáo khoa học Online của Đại học Cornell: <https://arxiv.org/>
- Towards Data Science Blog: <https://towardsdatascience.com/>
- Vũ Hữu Tiệp - Machine Learning Cơ Bản – 12/2018
- Andrew Ng - Machine Learning Course
- Duyệt - NLP - Truyện Kiều Word2vec – 2017:
<https://blog.duyet.net/2017/04/nlp-truyen-kiem-word2vec.html#.XE0oclwzaMq>
- Zalo Tech Blog – 11/1/2018: <https://medium.com/zalotech/sua-loi-chinh-ta-tieng-viet3bb0990af931>
- Adrian Rosebrock - PyImageSearch Gurus – 2018
- Vũ Hữu Tiệp - Machine Learning cơ bản – 2018
- Nguyễn Văn Long - Ứng dụng xử lý ảnh trong thực tế với thư viện OpenCV 05/2014
- Duy Luân - Viết engine phân loại hình ảnh bằng neural network một cách dễ dàng
- Phạm Hoàng Anh - Xây dựng chương trình tóm tắt văn bản (tiếng Việt) đơn giản với Machine Learning – 7/9/2019: <https://viblo.asia/p/xay-dung-chuong-trinh-tom-tat-van-ban-tieng-viet-don-gian-voi-machine-learning-YWOZrgAwlQ0>.
- Khanh Nguyen, Hieu Pham - Những Vấn đề Cơ bản về Mạng Neuron – 5/2/2016
- Các tài liệu sử dụng của các mã nguồn mở, API mà tác giả đề cập, sử dụng trong đề tài này.