

ỨNG DỤNG TEACHABLE MACHINE TRONG NHẬN DIỆN KHUÔN MẶT THEO THỜI GIAN THỰC

APPLICATION OF TEACHABLE MACHINE IN REAL-TIME FACE RECOGNITION

Trần Sinh Biên^{1*}, Đỗ Thị Chang¹

DOI: <https://doi.org/10.57001/huih5804.2023.010>

TÓM TẮT

Ngày nay, AI (Artificial Intelligence) và ML (Machine Learning) là hai trong những lĩnh vực ứng dụng hứa hẹn nhất trên thế giới. Tuy nhiên, với một người không chuyên công nghệ thông tin, rất khó để họ có thể tiếp cận và ứng dụng AI hay ML vào trong các nghiên cứu của mình. Để khắc phục vấn đề này, tác giả xin đề cập tới một công cụ do Google phát hành có tên Teachable Machine, Teachable Machine sẽ giúp chúng ta có được những trải nghiệm cơ bản nhất về trí thông minh nhân tạo (AI). Teachable Machines được xây dựng dựa trên mạng nơ-ron tích chập (CNN - Convolutional neural networks) - một trong những mô hình Deep Learning tiên tiến, nó giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao. Trong phạm vi bài báo này, tác giả cũng sẽ xây dựng một mô hình máy học dựa trên công cụ Teachable Machine nhằm xây dựng hệ thống nhận diện khuôn mặt trong thời gian thực. Bài báo sẽ tập trung vào vấn đề thu thập dữ liệu từ các đối tượng khác nhau, từ đó sử dụng công cụ Teachable Machine để xây dựng mô hình, và sau đó sẽ tiến hành nhận diện người trong thời gian thực thông qua camera. Sau 2 lần training mô hình, độ chính xác (accuracy) của mô hình đã đạt tới 100%.

Từ khóa: Học máy (ML), mạng nơ-ron tích chập (CNN), trí thông minh nhân tạo (AI).

ABSTRACT

These days, AI (Artificial Intelligence) and ML (Machine Learning) are two of the most promising application areas in the world. However, for a non-IT person, it is very difficult for them to approach and apply AI or ML in their research. To solve this problem, the author would like to mention a tool released by Google called Teachable Machine, Teachable Machine will help us get the most basic experience of artificial intelligence (AI). Teachable Machines is built based on a convolutional neural network (CNN - Convolutional neural networks) - one of the advanced Deep Learning models, it helps us build intelligent systems with high accuracy. Many users in 201 countries use Teachable Machine to create more than 125,000 classification models [2]. Within the scope of this paper, the author will also build a machine learning model based on the Teachable Machine tool to build a face recognition system in real-time. The article will focus on the problem of collecting data from different objects, then using the Teachable Machine tool to build a model, and then proceeding to identify people in real-time through the camera. After 2 times of training the model, the model's accuracy reached 100%.

Keywords: Machine learning (ML), Convolutional neural network (CNN), Artificial intelligence (AI).

¹Khoa Điện - Điện tử, Trường Đại học Hàng hải Việt Nam

*Email: transinhbien@vmaru.edu.vn

Ngày nhận bài: 05/01/2023

Ngày nhận bài sửa sau phản biện: 10/02/2023

Ngày chấp nhận đăng: 24/02/2023

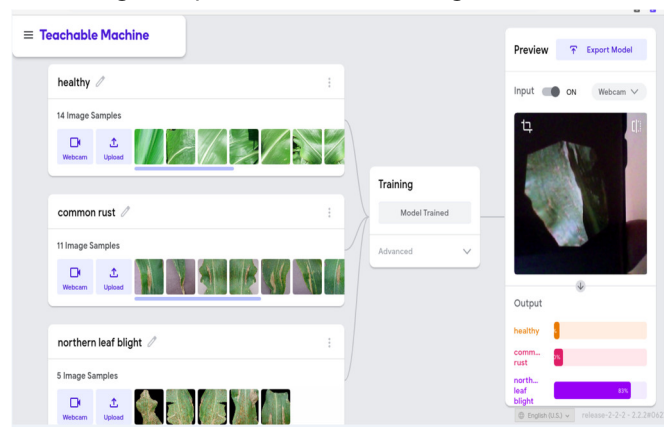
1. GIỚI THIỆU

Học máy (Machine learning) ngày càng phổ biến trong cuộc sống hàng ngày, cung cấp các xử lý vượt trội trong dịch thuật, phiên âm giọng nói và hơn thế nữa. Tuy nhiên, với những đối tượng chưa có nhiều trải nghiệm trong lĩnh vực công nghệ mới này, thậm chí là với những nhà nghiên cứu, nhà khoa học không có thể mạnh về mảng công nghệ thông tin, thì việc tạo ra một mô hình hoàn chỉnh cho những dự án của họ còn gặp rất nhiều khó khăn. Với vô vàn các kiến thức chuyên sâu, sự phát triển trong các thuật toán hiện đại, cũng như khả năng tác động tới xã hội của ML, thì điều cần thiết là cần cho phép những người ở nhiều lĩnh vực nghiên cứu khác-những người chưa biết nhiều về cách thức hoạt động và không có khả năng tạo ra mô hình của riêng họ. Một cách tiếp cận để mở rộng sự tương tác với ML là sử dụng các công cụ cho phép người dùng không có kiến thức kỹ thuật có thể tạo các mô hình ML của riêng họ. Một số môi trường máy học mã nguồn mở này là Google Colaboratory, Kaggle Kernal, đây là những nền tảng tuyệt vời cho các ứng dụng máy học và học sâu trên đám mây. Cả hai đều là sản phẩm của google và yêu cầu kiến thức về khoa học dữ liệu để phát triển và đào tạo các mô hình sử dụng chúng. Tuy nhiên, vài năm trở lại đây Google cũng đã giới thiệu một nền tảng mã nguồn mở mới để đào tạo các mô hình máy học đơn giản hơn rất nhiều so với các công cụ trước đó, đó là Teachable Machine. Teachable Machine được xây dựng bởi kỹ sư phần mềm Nikhil Thorat vào năm 2017 và được cải tiến thêm một số tính năng bổ sung trong phiên bản 2.0 vào năm 2019. Teachable Machine là một giao diện dựa trên web cho phép chúng ta đào tạo các mô hình phân loại ML mà không cần xây dựng những đoạn mã phức tạp. Teachable Machine có thể đào tạo và xây dựng các mô hình dựa trên tập dữ liệu hình ảnh, hình dáng cơ thể hoặc âm thanh, trong phạm vi bài báo này tôi chỉ đề cập tới vấn đề phân loại hình ảnh.

Hình 1 cho thấy ứng dụng được xây dựng bằng tiện ích mở rộng Teachable Machine để phân loại sâu bệnh cho cây ngô. Ứng dụng hiện đang dự đoán cây ngô đang bị ảnh hưởng bởi bệnh đốm lá, độ chính xác lên tới 83%.

Sự thu hút rộng rãi của Teachable Machine cho thấy nó đã trao quyền cho mọi người học, dạy và khám phá các khái niệm ML. Đã có rất nhiều người dùng trên 201 quốc gia ứng dụng Teachable Machine để tạo ra hơn 125.000 mô hình

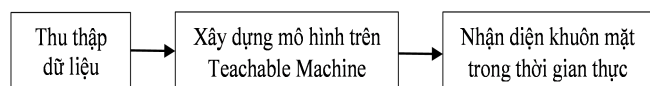
phân loại [2]. Mọi người đã tạo ra chương trình giảng dạy, hướng dẫn và các tài nguyên khác bằng Teachable Machine về các chủ đề như liên quan tới AI tại các tổ chức như Học viện Thiết kế Stanford, Chương trình Viễn thông Tương tác của NYU, MIT Media Lab [2], chuyên gia công nghệ Steve Sailing đã xây dựng một công cụ giao tiếp cho những người mắc chứng nói lắp [3], và nhiều các thử nghiệm khác.



Hình 1. Ứng dụng Teachable Machine trong phân loại sâu bệnh gây hại cho thực vật [1]

Như vậy, bằng việc ứng dụng Teachable Machine chúng ta hoàn toàn có thể xây dựng những mô hình học máy để ứng dụng trong các hệ thống phân loại sản phẩm theo hình dạng, điều khiển Robot bằng giọng nói hoặc cử chỉ,... Trong phạm vi bài báo này, tác giả sẽ sử dụng Teachable Machine để đào tạo và xây dựng mô hình học máy ứng dụng trong hệ thống nhận diện khuôn mặt theo thời gian thực. Công nghệ Camera nhận diện khuôn mặt là công nghệ sinh trắc học ít tác động đến người dùng nhất và là công nghệ sinh trắc học nhanh nhất. Hệ thống sẽ nhận diện khuôn mặt một cách kín đáo bằng cách chụp ảnh khuôn mặt của những ai bước vào một khu vực được xác định từ camera giám sát. Trong cuộc sống hiện đại, công nghệ AI nhận diện khuôn mặt có tác động không hề nhỏ đối với các hoạt động của đời sống con người. Công nghệ này giúp cho hoạt động của con người trở nên tiện ích, hiệu quả hơn.

Cấu trúc chương trình gồm 3 bước được thể hiện trên hình 2.



Hình 2. Cấu trúc chương trình

Bước 1: Thu thập dữ liệu: Sử dụng mã nguồn mở OpenCV để phát hiện, lưu trữ 500 bức ảnh của từng đối tượng trong từng file riêng biệt

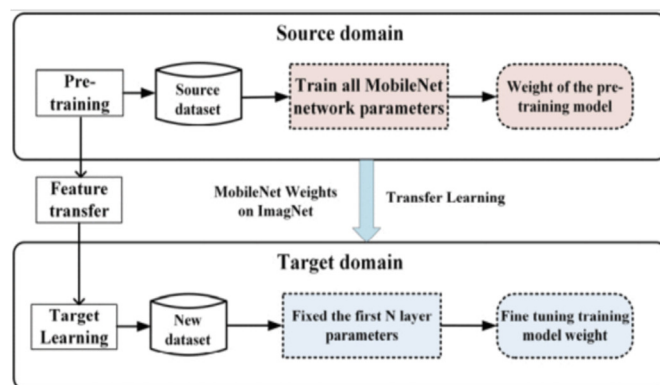
Bước 2: Sử dụng nguồn dữ liệu đã thu thập được ở bước 1 để huấn luyện và tải mô hình sử dụng công cụ Teachable Machine.

Bước 3: Nhận diện khuôn mặt. Bước này sẽ tiến hành so sánh dữ liệu được trích xuất từ camera (sau khi đã được xử lý dữ liệu đầu vào) với dữ liệu đã được huấn luyện để trích xuất kết quả lên ngay màn hình quan sát.

2. HUẤN LUYỆN MÔ HÌNH SỬ DỤNG CÔNG CỤ TEACHABLE MACHINE

2.1. Giới thiệu công cụ Teachable Machine

Teachable Machine sử dụng phương pháp học chuyển giao, một kỹ thuật ML để tìm các mẫu và xu hướng trong hình ảnh hoặc mẫu âm thanh, đồng thời tạo ra một mô hình phân loại đơn giản và dễ dàng trong vòng vài giây. Để xây dựng các lớp phân loại hình ảnh, Teachable Machine dựa trên mạng nhận dạng hình ảnh được đào tạo trước gọi là MobileNet - Mạng này đã được đào tạo để nhận ra hàng ngàn đối tượng (chẳng hạn như mèo, chó, ô tô, trái cây và chim). Hình 3 cho thấy quá trình đào tạo TL-MobileNet dựa trên mô hình học chuyển giao. Các trọng số và đặc trưng của mô hình MobileNet được đào tạo trước trong bộ miền nguồn (Source domain), sau đó chúng được chuyển sang miền đích (Target domain) để phân loại. Miền đích không sử dụng khởi tạo ngẫu nhiên để bắt đầu quá trình học dữ liệu ngay từ đầu và các tham số mô hình được chia sẻ giữa miền nguồn và miền đích, vì vậy phương pháp này sẽ giúp cải thiện hiệu quả học tập.



Hình 3. TL-MobileNet dựa trên mô hình học chuyển giao. [4]

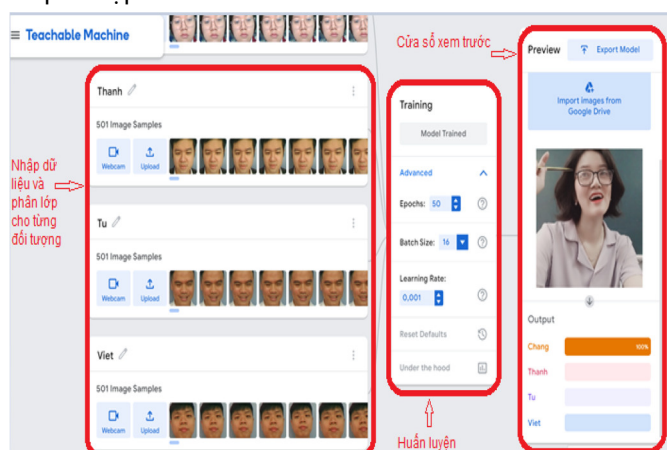
Với Teachable Machine chúng ta có thể tạo các mẫu dữ liệu trực tiếp từ webcam hoặc các dữ liệu đã có từ trước đó để đào tạo mô hình. Công cụ này của Google sử dụng thuật toán KNN (K-Nearest Neighbors) [5] một trong những thuật toán học máy có giám sát, đơn giản và dễ triển khai - để phân loại hình ảnh. Giao diện Teachable Machine mà chúng ta sử dụng chỉ là front-end do Google tạo ra, nó không bao gồm bất cứ thông tin mã nguồn nào về Back-end, hay nói cách khác thông tin về các lớp mạng ẩn không được Google công bố nên chúng ta không thể biết được về chúng.

2.2. Huấn luyện và trích xuất mô hình nhận diện khuôn mặt

Có ba bước liên quan đến việc đào tạo mô hình với Teachable Machine - thu thập mẫu, đào tạo mô hình, thử nghiệm và cải tiến mô hình. Số lượng mẫu cần đủ lớn để phục vụ cho quá trình huấn luyện. Quá trình đào tạo sẽ diễn ra ngay trên trình duyệt. Dữ liệu vẫn sẽ nằm trong máy của chúng ta và sẽ không bị tải lên đám mây, nên sẽ đảm bảo tính bảo mật cho nguồn dữ liệu đầu vào. Cuối cùng, chúng ta có thể điều chỉnh mô hình để cải thiện độ chính xác cho mô hình. Ngoài ra, chúng ta còn có thể bổ sung mẫu và huấn luyện lại nhằm tăng độ chính xác cho dự đoán.

Một trong những thế mạnh của Teachable Machine là nó có khả năng xuất mô hình sang các môi trường và ứng dụng khác nhau. Chúng ta có thể dễ dàng xuất mô hình dưới dạng mô hình TensorFlow.js và lưu trữ miễn phí trên Teachable Machine để có thể sử dụng mô hình đó với bất kỳ trang web hoặc ứng dụng dành cho thiết bị di động nào. Cũng có thể chuyển đổi nó thành TensorFlow và TensorFlow Lite và tải xuống để sử dụng cục bộ. Trong dự án này, tác giả đã chuyển đổi thành TensorFlow và lưu trữ về máy để phục vụ cho chương trình nhận diện khuôn mặt sau này.

Để thực hiện đào tạo mô hình, chúng ta cần thay đổi một số thông số của mô hình trong cửa sổ Training - cửa sổ huấn luyện (hình 4) như: epoch - 1 epoch là một lần duyệt qua hết các dữ liệu trong tập huấn luyện; Batch Size - dữ liệu cho mỗi lần tính và cập nhật hệ số; Learning rate - Tốc độ học sẽ kiểm soát tốc độ mô hình thay đổi các trọng số để phù hợp với bài toán.



Hình 4. Quá trình huấn luyện mô hình trên Teachable Machine

Trong phạm vi bài báo, tác giả thực hiện thu thập dữ liệu của 4 đối tượng, mỗi đối tượng thu thập 500 bức ảnh khuôn mặt. Sau đó, trên Teachable Machine, tác giả phân lớp cho từng đối tượng và thực hiện huấn luyện với thông số: 50 epochs; Batch Size là 16, Learning Rate - 0.001. Hình 4 thể hiện quá trình huấn luyện mô hình trên Teachable Machine.

3. THUẬT TOÁN CỦA CHƯƠNG TRÌNH

3.1. Thuật toán thu thập dữ liệu

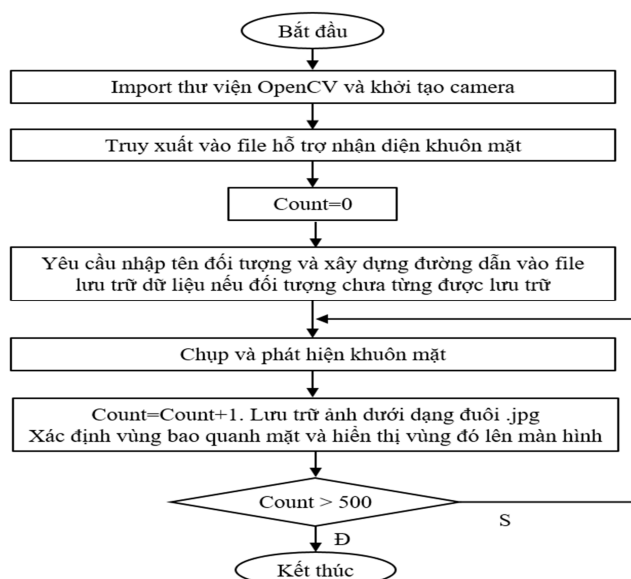
Để thu thập dữ liệu, thay vì lấy nguồn dữ liệu trên mạng, tác giả đã thực hiện thu thập bằng chương trình được viết trên ngôn ngữ python, sử dụng thư viện mã nguồn mở để phát hiện khuôn mặt người. Hình 5 thể hiện thuật toán thu thập dữ liệu.

Dữ liệu sau khi thu thập dưới dạng file ảnh .jpg sẽ được lưu trữ ở mỗi file riêng biệt cho từng đối tượng, với mỗi đối tượng sẽ thu thập 500 file ảnh. Mỗi file dữ liệu này sau đó sẽ được chia làm hai phần tự động bằng công cụ Teachable Machine để phục vụ cho quá trình đào tạo và kiểm tra độ chính xác của mô hình sau này:

- Tập dữ liệu phục vụ quá trình thử nghiệm cuối cùng - Testing set (chiếm 15%);

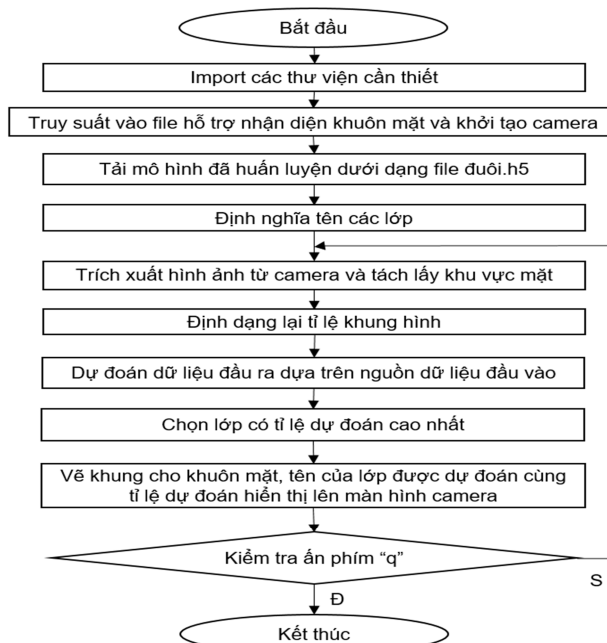
- Tập dữ liệu phục vụ quá trình huấn luyện (chiếm 85%).

Trong quá trình thu thập dữ liệu, điều kiện về ánh sáng, nền ảnh đằng sau,...cũng làm ảnh hưởng tới chất lượng tạo mô hình sau này.



Hình 5. Thuật toán thu thập dữ liệu

3.2. Thuật toán nhận diện khuôn mặt theo thời gian thực



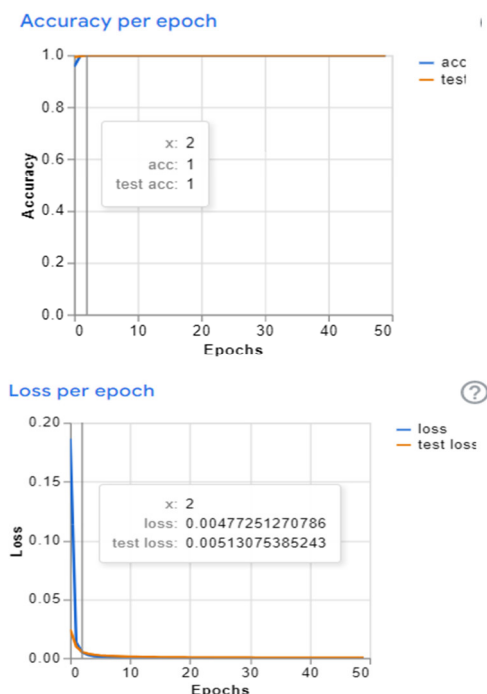
Hình 6. Thuật toán nhận diện khuôn mặt theo thời gian thực

Để phát hiện khuôn mặt, nhóm tác giả sử dụng thuật toán phát hiện khuôn mặt Viola-Jones [6] giống như trong chương trình thu thập dữ liệu.

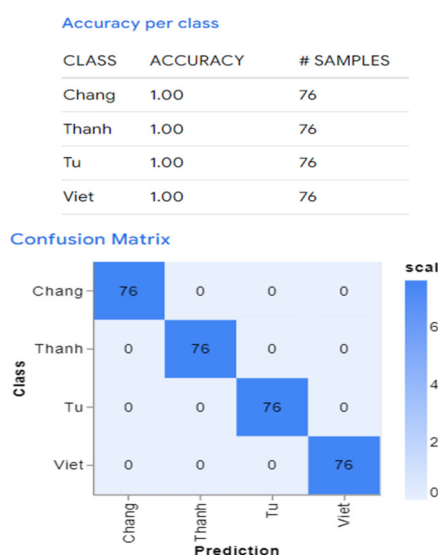
Sau khi ảnh được cắt từ video, nó sẽ được định dạng lại kích thước giống với kích thước của dữ liệu trong mô hình huấn luyện. Khi đó các tham số sinh ra được từ tập dữ liệu huấn luyện sẽ được sử dụng để thẩm định lại tính thích hợp

của mô hình trên tập dữ liệu của hình ảnh vừa được trích xuất. Dựa vào kết quả thu được ta tiến hành hiển thị lên màn hình kết quả đối tượng có tỉ lệ dự đoán cao nhất, cùng với tên và tỉ lệ dự đoán của đối tượng. Thuật toán xây dựng như trong hình 6.

4. KẾT QUẢ CHẠY THỬ NGHIỆM



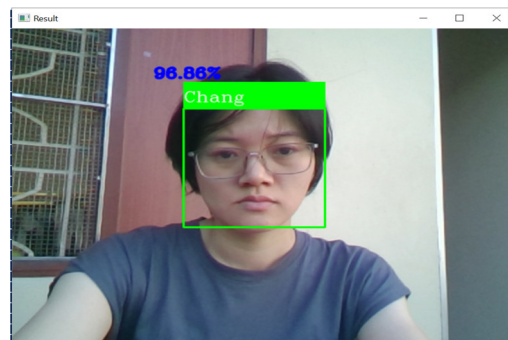
Hình 7. Độ chính xác và độ mất mát của mô hình



Hình 8. Độ chính xác trên từng lớp và ma trận lỗi của mô hình

Sau 2 lần training mô hình, độ chính xác (accuracy) của mô hình đã đạt tới 100%, độ mất mát (loss) gần như về 0%. Như vậy, ta hoàn không cần tới 50 epochs trong quá trình đào tạo mô hình, từ đó giúp giảm thời gian đào tạo mô hình. Hình 7 cho thấy dạng đồ thị của độ chính xác và sự mất mát của mô hình; Độ chính xác trên mỗi lớp và ma trận lỗi được thể hiện trên hình 8.

Chương trình thực hiện nhận diện khá chính xác, trong một vài thử nghiệm có tỉ lệ dự đoán lên tới 100%. Tuy nhiên, khi điều kiện ánh sáng kém việc nhận diện vẫn chưa thực sự hiệu quả. Khoảng cách của đối tượng đến camera cũng là một yếu tố ảnh hưởng lớn tới chất lượng nhận diện khuôn mặt.



Hình 9. Kết quả nhận diện trong thời gian thực

5. KẾT LUẬN

Bài báo đã xây dựng thành công chương trình nhận diện khuôn mặt trong thời gian thực ứng dụng Teachable Machine. Chương trình được viết bằng ngôn ngữ Python và sử dụng một số thư viện mã nguồn mở như OpenCV, Tensorflow, Keras,... Dựa vào những kết quả thu được, ta có thể kết hợp chương trình với các thiết bị phần cứng như Raspberry, Arduino,... để xây dựng hệ thống giám sát, mở khóa bằng khuôn mặt,...

LỜI CẢM ƠN

Nghiên cứu này được tài trợ bởi Trường Đại học Hàng hải Việt Nam trong đề tài mã số: DT22-23.58.

TÀI LIỆU THAM KHẢO

- [1]. Kimaru Thagana, 2020. *Getting Started with Image Classification Using Google Teachable Machine*. <https://www.pluralsight.com/guides/getting-started-with-google-teachable-machine>.
- [2]. Michelle Carney, Barron Webster, Irene Alvarado, Kyle Phillips, Noura Howell, Jordan Griffith, Jonas Jongejan, Amit Pitaru, Alexander Chen, 2020. *Teachable Machine: Approachable Web-Based Tool for Exploring Machine Learning Classification*. CHI 2020, Honolulu, HI, USA.
- [3]. Sumbo Bello, 2019. *Teachable Machine Gets New Features For Training AI Models*. <https://edgy.app/teachable-machine-gets-new-features-for-training-ai-models>
- [4]. H. Pan, Z. Pang, Y. Wang, Y. Wang, L. Chen, 2020. *A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects*. in IEEE Access, vol. 8, pp. 119951-119960, doi: 10.1109/ACCESS.2020.3005450.
- [5]. *Teachable-machine-boilerplate*. <https://github.com/googlecreativelab/teachable-machine-boilerplate>, 2018.
- [6]. P. Viola, M. J. Jones, 2004. *Robust real-time face detection*. Int. J. Comput. Vision, Vol. 57, No. 2, pp. 137-154.

AUTHORS INFORMATION

Tran Sinh Bien, Do Thi Chang

Faculty of Electrical - Electronic Engineering, Vietnam Maritime University