**XÂY DỰNG HỆ THỐNG NHẬN DẠNG**

**ĐỐI TƯỢNG SỬ DỤNG KHẨU TRANG**

**Lê Tấn Lộc, Huỳnh Thanh Phong,**

**Trần Thị Tứ, Linh, Phạm Duy Minh**

(Sinh viên năm 2, Khoa *Công Nghệ Thông Tin*)

GVHD*: ThS. Lương Trần Ngọc Khiết*

**Tóm tắt**: Convolutional Neural Network (CNN – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến. Nó giúp cho chúng ta xây dựng được những hệ thống thông minh để nhận dạng được các đối tượng và đưa ra kết quả với độ chính xác cao như hiện nay. Mục tiêu của đề tài tập trung vào bài toán nhận diện những người có đeo khẩu trang đảm bảo hạn chế sự lây lan của dịch bệnh, từ đó xây dựng một chương trình có hiệu quả, đầy đủ và chính xác với mục đích hỗ trợ việc nhận dạng thực thể song xây dựng một hệ thống huấn luyện, thực nghiệm và ứng dụng.

**Từ khóa**: Convolutional Neural Network, Deep Learning, Face Mask Detection, MobileNetV2, COVID-19

# Mở đầu

## Giới thiệu chung

Hiện nay trong thời kỳ COVID-19 đang diễn biến phức tạp khiến cho bài toán quan sát mọi người có đeo khẩu trang hay không đóng vai trò quan trọng trong việc phòng chống chiến đấu với COVID-19. Mô hình mạng nơ-ron tích chập (Convolutional Neural Network - CNN) là một mô hình tiêu biểu để có thể giải quyết bài toán này.

Việc đề xuất một mô hình nhận dạng đối tượng đeo khẩu trang có độ chính xác cao là nhu cầu tất yếu trong thực tế và hiện đang được đông đảo các nhà nghiên cứu quan tâm. Đặc biệt đối với các nước đang có sự bùng nổ nghiêm trọng ảnh hưởng trực tiếp đến thế giới

## Bài toán nhận diện

Nói về mảng Khoa học máy tính nó có thể thực hiện việc tái tạo các bộ phận phức tạp của hệ thống thị giác của con người, cho phép máy tính xác định và xử lý các đối tượng trong hình ảnh và video giống cách con người làm.

Nhờ những tiến bộ về trí tuệ nhân tạo và cải cách trong việc nghiên cứu chuyên sâu và mạng lưới thần kinh, lĩnh vực này đã có những bước phát triển vượt bậc.

Trong số những yếu tố thúc đẩy sự phát triển của thị giác máy tính là lượng dữ liệu chúng ta tạo ra được sử dụng để đào tạo và cải thiện hệ thống cảm biến thị giác máy tính, nó đóng vai trò quan trọng trong các ứng dụng nhận dạng khuôn mặt.

Qua thuật toán thị giác máy tính phát hiện các đặc điểm khuôn mặt trong hình ảnh và so sánh chúng với cơ sở dữ liều về cấu hình khuôn mặt. Từ đó có thể nhận diện nhanh danh tính của đối tượng thông qua khuôn mặt.

Định nghĩa bài toán nhận diện

* Bước 1: Cho dữ liệu đầu vào là một hình ảnh.
* Bước 2: Phân tích hình ảnh để xác định những vùng có khuôn mặt
* Bước 3: Kiểm tra việc đeo khẩu trang ở những vùng có khuôn mặt
* Bước 4: Đưa ra kết luận

|  |  |  |
| --- | --- | --- |
| A picture containing person, person, posing  Description automatically generated | A person wearing glasses  Description automatically generated with medium confidence | 🡺Không Khẩu Trang |
| A picture containing person  Description automatically generated | 🡺Có Khẩu Trang |

Hình 1‑1 Mô phỏng bài toán.

Từ đó có thể được áp dụng trên các thiết bị trực tiếp (kiểm tra trên thời gian thực) để kiểm tra việc đối tượng sử dụng khẩu trang ngoài thực nghiệm.

## Lý do chọn đề tài

Convolutional neural networks đã được sử dụng nhiều trong các bài toán nhận dạng vật thể. Tuy nhiên để áp dụng một bài toán cụ thể ta cần xây dựng mô hình huấn luyện với số lượng lớp và mỗi lớp là một bài toán phân tích.

Đề xuất mô hình mạng CNN theo giải thuật tối ưu để nhận dạng các vật thể và đưa ra kết luận có phải vật thể đó hay không. Đặc biệt là trong giai đoạn dịch COVID-19 như hiện nay thì việc nhận dạng đối tượng sử dụng khẩu trang là vấn đề nóng được nhiều nước trên thế giới quan tâm. sức ảnh hưởng đến thị trường kinh tế.

# Mục đích nghiên cứu, đối tượng nghiên cứu, phạm vi nghiên cứu, phương pháp nghiên cứu

## Mục đích nghiên cứu

Ở nghiên cứu này tập trung vào bài toán nhận diện những người có đeo khẩu trang đảm bảo hạn chế sự lây lan của dịch bệnh. Mục tiêu cụ thể của đề tài là phân tích, phát triển vấn đề, đề xuất giải pháp cho việc nhận diện các loại đối tượng có đeo khẩu trang thuộc miền dữ liệu trên. Xây dựng một chương trình có hiệu quả, đầy đủ và chính xác với mục đích hỗ trợ việc nhận dạng thực thể song xây dựng một hệ thống huấn luyện, thực nghiệm và ứng dụng.

## Đối tượng

Đề tài lựa chọn đối tượng đầu vào để nghiên cứu gồm

* Bài toán nhận diện.
  + Nhận diện khuôn mặt
  + Nhận diện khuôn mặt đeo khẩu trang
* Dữ liệu hình ảnh.
  + Thống kê dữ liệu
  + Phân tích dữ liệu
* Công cụ máy học CNN.
  + Phân tích các điểm ảnh: Mô hình mạng nơ-ron thần kinh tích chập (convolutional neural networks - CNN)
  + Nhận dạng các điểm ảnh đặc trung của hình ảnh: Phương pháp tính chập mạng nơ-ron (CNNs)

## Phạm vi nghiên cứu

Phạm vi nghiên cứu của đề tài là nhận diện được đối tượng có đeo khẩu trang hay không trong video thời gian thực (ghi hình trực tiếp) trên các thiết bị ghi hình.

Phạm vi áp dụng có thể áp dụng trong các siêu thị, cửa hàng tiện lợi, tiệm tạp hóa, trường học, trung tâm thương mại… Đối tượng được xử lý phải đảm bảo có 2/3 khuôn mặt được xác định qua khung hình camera thì mới có thể nhận diện được.

## Phương pháp nghiên cứu

Các phương pháp nghiên cứu được áp dụng nhằm xác định kết quả gồm:

* Mô hình hóa đề xuất mô hình máy học dựa trên CNN cho bài toán nhận diện đối tượng đeo khẩu trang ở những nơi công cộng đông người.
* Cài đặt thực nghiệm trên máy tính.
* Phân tích hình ảnh, đánh giá và đưa kết luận.

# Kết quả nghiên cứu

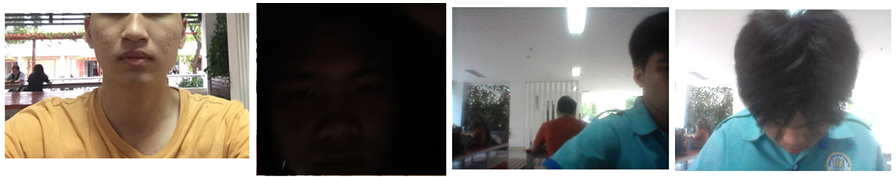
## Cơ sở

* **Quy hoạch dữ liệu đầu vào**:

Thu thập dữ liệu phải đảm bảo các yêu cầu sau: phải có trên 2/3 khuôn mặt.trong điều kiện phải đủ sáng, không bị ngược sáng và chất lượng ảnh tối thiểu phải từ 360p trở lên.



Hình 3‑1 Một số dữ liệu đạt yêu cầu.



Hình 3‑2 Một số dữ liệu không đạt yêu cầu

* Thu thập dữ liệu: gắn nhãn dữ liệu thủ công, tổ chức thành cây thư mục
* Bộ dữ liệu gồm 2 nguồn được tổ hợp chung lại với nhau là Sriman\_Mitra và bộ dữ liệu tự tạo từ các thành viên trong nhóm. Do yêu cầu đầu ra là kết quả “Có” hoặc “Không” nên gắn nhãn dữ liệu thủ công bằng cách phân chia 2 thư mục lưu trữ là with\_mask; without\_mask được lưu trong thư mục dataset.

|  |  |
| --- | --- |
|  |  |

* Với quy ước gắn nhãn cho ảnh: Attribute\_Status(nếu có)\_Index. Trong đó:
* ***Attribute:*** nhận giá trị *mask* nếu có khẩu trang, hoặc nhận giá trị *notmak* nếu không khẩu trang.
* ***Status:*** nhận giá trị *multi* nếu trong ảnh có nhiều người.
* ***Index:*** là số thứ tự ảnh với 4 số.

|  |  |
| --- | --- |
| Thư mục with\_mask | Thư mục without\_mask |
| mask\_0003.jpg | notmask\_0003.jpg |
|  |  |
|  |  |
| Hình 3‑3 Bộ dữ liệu đối tượng sử dụng khẩu trang | Hình 3‑4 Bộ dữ liệu đối tượng không sử dụng khẩu trang |

* Phân chia tập dữ liệu thành tập huấn luyện, kiểm tra và kiểm định.

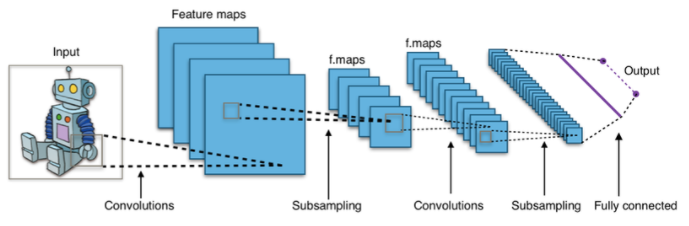
|  |  |
| --- | --- |
| Bảng 3‑1 Thống kê số lượng hình ảnh training và testing. | Bảng 3‑2 Các tham số trong mô hình. |
| |  |  |  | | --- | --- | --- | |  | Training (80%) | Testing (20%) | | Có khẩu trang | 1647 | 412 | | Không khẩu trang | 1598 | 399 | | |  |  | | --- | --- | | **Tham số** | **Giá trị** | | INIT\_LR |  | | EPOCHS | 20 | | BS | 32 | |

* **Mạng CNN**

Mạng nơ-ron tích chập (CNN) là một lĩnh vực nghiên cứu rất lớn trong lĩnh vực trí tuệ nhân tạo, CNN được xem như một hệ thống kết nối tập hợp các ngõ vào (inputs) đến tập hợp các ngõ ra (outputs) qua một hay nhiều lớp nơ-ron, các lớp này được gọi là các lớp ẩn.

Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo. Các lớp liên kết được với nhau thông qua cơ chế convolution, lớp tiếp theo là kết quả convolution từ lớp trước đó. Như vậy, mỗi neuron ở lớp kế tiếp sinh ra từ kết quả của filter áp đặt lên một vùng ảnh cục bộ của neuron trước đó. Mỗi một lớp được sử dụng các filter khác nhau thông thường có hàng trăm hàng nghìn filter như vậy và kết hợp kết quả của chúng lại.

Ngoài ra có một số layer khác như pooling/subsampling layer dùng để chắt lọc lại các thông tin hữu ích hơn (loại bỏ các thông tin nhiễu).



Hình 3‑5 Mô hình về cấu trúc mạng Convolution Neural Network - CNN.

* **Các thư viện sử dụng.**

Các thư viện chính.

* ***TensorFlow*** là một thư viện mã nguồn mở để tính toán số và học máy quy mô lớn. TensorFlow kết hợp một loạt các mô hình và thuật toán học máy và học sâu (hay còn gọi là mạng Neural) và làm cho chúng trở nên hữu ích bằng một phép ẩn dụ thông thường. Dùng để làm nền tảng cho việc sử dụng thư viện Keras.
* ***Keras*** là một thư viện cung cấp một API rất dễ sử dụng cũng như đủ trực quan chạy với nền tảng là Tensorflow. Và đối với đề tài này Keras sẽ tập trung trong việc xử lý hình ảnh hay khung hình của máy quay thành một mảng numpy cho việc bỏ vào model và cũng như để nạp model nhận diện đối tượng khẩu trang.
* ***OpenCV-Python*** là một thư viện xử lý hình ảnh, dùng để đọc model cho nhận dạng khuôn măt và vẽ khung xác định những khuôn mặt và ghi kết quả nhận định trên ảnh hoặc các khung hình của máy quay trên thời gian thực. Sau đó sẽ hiển thị lên màn hình

Các thư viện bổ trợ.

* ***Numpy*** hỗ trợ cho việc tính toán các mảng nhiều chiều, có kích thước lớn..
* ***Matplotlib*** thư viện vẽ đồ thị.

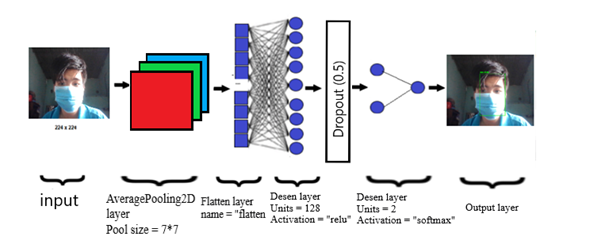
## Mô hình bài toán nhận dạng

### Mô hình

* **Cấu trúc mạng**

Đề tài sử dụng mô hình MobileNetV2

MobileNetV2 là một phiên bảng phát triển hơn của mô hình MobileNetV1 áp dụng cho việc xây dựng những ứng dụng sử dụng thị giác máy trên thiết bị điện thoại di động.



Hình 3‑6 Mô hình huấn luyện của đề tài.

Trong đó:

* AveragePooling2D layer: là lớp sẽ gộp những chi tiết của quan trọng thông qua bài toán tính chập hai chiều từ ảnh có kích thước *224* x *224* x *3* pixel xuống còn 32 x 32 x 3 pixel (Pool size = 7 \* 7 tức là gộp những chi tiết quan trọng của hỉnh ảnh sao cho kích thước chiều rộng và chiều cao giảm gấp 7 lần).
* Flatten layer: là lớp chuyển đổi dữ liệu ở lớp AveragePooling2D layer thành một vector một chiều, để nhập dữ liệu vào các lớp tiếp theo.
* Desen layer: là lớp mạng nơ-ron nhân tạo trong đó mỗi nơ-ron nhận đầu vào từ tất cả nơ-ron của lớp trước đó.
  + Units là số chiều đầu ra của lớp.
  + Activation là tham số kích hoạt (hàm tính toán) trong việc áp chức năng kích hoạt các phần tử trong vùng thỏa điều kiện tham số kích hoạt.

|  |  |
| --- | --- |
|  | . |
| Hình 3‑7 Đồ thị hàm tham số ReLu | Hình 3‑8 Đồ thị hàm tham số Softmax |

* Output layer: lớp sẽ hiển thị kết quả, với đối tượng có sử dụng khẩu trang sẽ hiển thị khung màu xanh, ngược lại là khung màu đỏ cùng với tỉ lệ nhận dạng tương ứng.

### Kết quả.

|  |  |
| --- | --- |
| Chart, line chart  Description automatically generated | Bảng ‑ Kết quả huấn luyện. |
| |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | Precision | Recall | F1-score | Support | | With\_mask | 0.98 | 0.99 | 0.99 | 412 | | Without\_mask | 0.99 | 0.98 | 0.99 | 399 | | Accuracy |  |  | 0.99 |  | |
| Hình ‑ Đồ thị huấn luyện. |

Trong đó:

*Đồ thị huấn luyện (Hình 3-9):*

train\_loss: dự đoán sai khi huấn luyện.

train\_acc: dự đoán đúng khi huấn luyện.

test\_loss: dự đoán sai khi kiểm tra.

test\_acc: dự đoán đúng khi kiểm tra.

Khi quan sát kết quả của bộ đo training cho thấy đường train\_acc và test\_acc gần tiến về 1.0 đồng thời train\_lost và test\_lost cũng gần tiến về 0.0, nghĩa là kết quả của việc nhận diện khuôn mặt đeo khẩu trang gấn đúng 100%.

Ngoài ra, từ kết quả ta nhận thấy kể từ epoch thứ 17 trở đi thì đường train\_acc và test\_acc trùng nhau tại vị trí 1.0 đồng thời đường train\_loss và test\_loss cũng trùng nhau tại vị trí 0.0, như vậy với mô hình này thì chạy epoch bằng 17 là đủ.

Hệ thống đã hoạt động trong nhiều khung cảnh khác nhau và phát hiện được các đối tượng dạng đơn (một người) và dạng phức (nhiều người) có mang khẩu trang hay không trên hình ảnh và cả trên real-time video.

*Kết quả huấn luyện (Bảng 3-3).*

*Support* là tập dữ liệu test với dòng thứ nhất có 412 bức ảnh được gán nhãn là with\_mask (trong hình ảnh đối tượng có đeo khẩu trang), dòng thứ hai có 399 bức ảnh được gán nhãn là without\_mask (trong hình ảnh đối tượng không đeo khẩu trang).

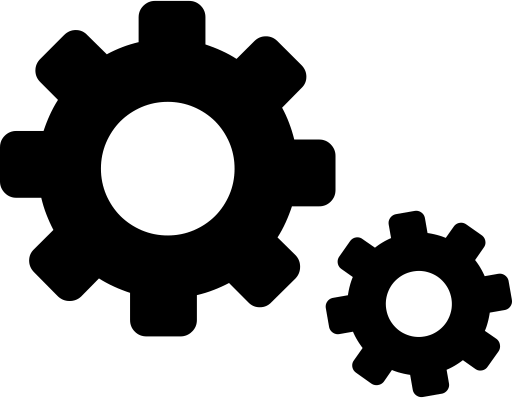
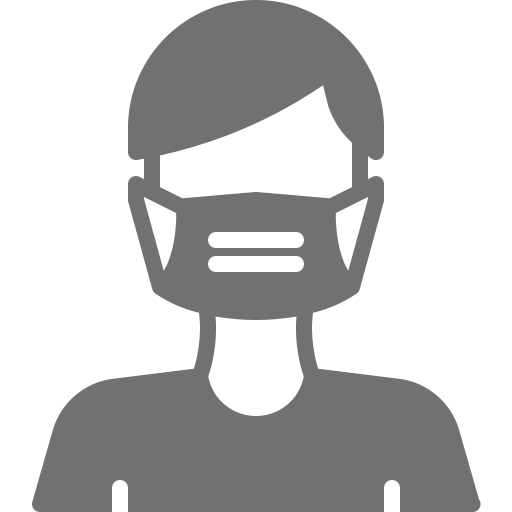
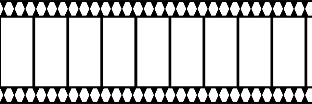
Giá trị của *Precision* ở dòng thứ nhất cho biết hệ thống dự doán bức ảnh đó gán nhãn là with\_mask (có đeo khẩu trang) là đúng đến 98%, hay nói cách khác là chỉ có 2% hệ thống dự đoán dự đoán nhầm là without\_mask (không có đeo khẩu trang). Trong khi bức ảnh đó gán nhãn là with\_mask (có đeo khẩu trang). Tương tự với dòng thứ hai.

Giá trị của *Recall* ở dòng thứ nhất cho biết hệ thống có thể nhận dạng được tới 99% trong số 412 bức ảnh gán nhãn là with\_mask là có khẩu trang. Tương tự ở dòng thứ hai.

Giá trị của *F1-score* cho biết mức độ hiệu quả của hệ thống đối với việc nhận dạng đối tượng có sử dụng khẩu trang ở dòng thứ nhất và nhận đạng đối tưởng không sử dụng khẩu trnag ở dòng thứ hai.

Giá trị của *Accuracy* cho biết mức độ hiệu quả của toàn bộ hệ thống đối với việc nhận đối tượng sử dụng khẩu trang.

Cách nhận dạng đối tượng khẩu trang trên real-time video:



Hình 3‑10 Minh họa ứng dụng nhận dạng đối tượng trong hệ thống camera giám sát

* Kết nối chương trình với camera của thiết bị và chạy trên thời gian thực
* Chương trình sẽ bắt từng frame từ camera
* Chỉnh sửa frame để phù hợp để bỏ vào hệ thống nhận diện
* Hệ thống sẽ trả lại và hiện frame sau khi đã qua hệ thống nhận diện với khung viền đỏ tương ứng với không mang khẩu trang, khung viền xanh với có mang khẩu trang và tỉ lệ chính xác của hệ thống cùng với dòng thông báo cùng màu trên góc trái trên của viền.

Một số hình ảnh thực tế từ hệ thống

|  |  |
| --- | --- |
| Hình 3‑11 Đối tượng không sử dụng khẩu trang. | Hình 3‑12 Đối tượng sử dụng khẩu trang. |
| Hình 3‑13 Một đối tượng sử dụng khẩu trang và một đối tượng không sử dụng khẩu trang. | Hình 3‑14 Thực nghiệm kiểm tra đeo khẩu trang cùng một lúc nhiều đối tượng. |

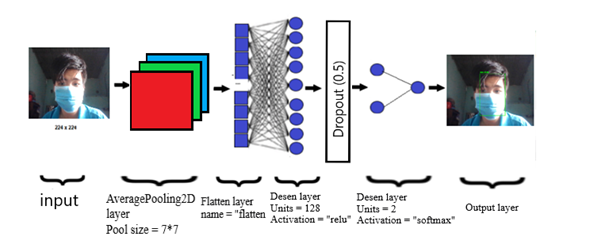
# Kết luận và hướng phát triển

## Kết luận

Nghiên cứu đã trình bày về vấn đề nhận dạng đối tượng sử dụng khẩu trang, một bài toán quan trọng trong lĩnh vực xử lý hình ảnh. Luận án tập trung nghiên cứu, phát triển về lý thuyết và ứng dụng đối với bài toán nhận dạng đối tượng, đề xuất một số mô hình và giải pháp nhằm nâng cao hiệu quả nhận dạng đối tượng sử dụng khẩu trang và đưa ra một số khung làm việc phục vụ cho quá trình nhận dạng đối tượng.

Nâng cao chất lượng nhận dạng đối tượng liên quan tới khuôn mặt sử dụng khẩu trang bằng cách thu nhập số lượng dữ liệu hình ảnh đủ lớn, tiến hành thực nghiệm ngoài thực tế để ghi chép những những sai sót, lập bảng thống kê về các tham số để chương trình được tối ưu hơn.

Tuy nhiên, hệ thống trong đề tài mà chúng em đang nghiên cứu vẫn còn một số hạn chế như không thể nhận diện hoặc nhân diện sai trong điều kiện thiếu ánh sáng, hay là dùng một vật thể khác mà không phải là khẩu trang che khuôn mặt. Mặc dù nghiên cứu còn những hạn chế và thiếu sót nhất định, nhưng chúng em sẽ tiếp tục cố gắng hoàn thiện trong thời gian tới.



Hình 4‑1 Mô hình minh họa thiết kê model.

## Hướng phát triển

Hiện tại, do hạn chế về thời gian, nghiên cứu dừng lại thử nghiệm với tập dữ liệu có kích cỡ vừa. Sau này chúng em sẽ mở rộng thử nghiệm trên tập dữ liệu lớn hơn.

Mở rộng và thử nghiệm trên các kiểu thực thể và mối quan hệ thực thể khác. Cải tiến áp dụng các dạng biểu đồ trực quan khác nhau trong việc phân tích các thực thể.

Sau khi trải qua mùa dich COVID thì hệ thống vẫn có thể đổi lại cài đặt giống như lúc trước dịch. Vì việc đeo khẩu trang lúc này không còn bắt buộc nữa.

Ngoài ra, nhóm chúng em còn sẽ có định hướng đến việc kết hợp hệ thống nhận diện khẩu trang với một hệ thống nhận biết danh tính của một người đó trên không gian thực, nhằm mục đích là có thể nhận biết danh tính của một người, kể cả khi họ đang đeo khẩu trang, với hệ thống này có thể được áp dụng trong hệ thống điểm danh ở các công ty, ...

# TÀI LIỆU THAM KHẢO

* Nguồn thông tin tham khảo

1. Mạng nơ ron tích chập cheatsheet: Đại học Stanford

(<https://stanford.edu/~shervine/l/vi/teaching/cs-230/cheatsheet-convolutional-neural-networks>)

1. S. A. Sanjaya and S. Adi Rakhmawan, "Face Mask Detection Using MobileNetV2 in The Era of COVID-19 Pandemic," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), Sakheer, Bahrain, 2020, pp. 1-5,
2. Valueva, M.V.; Nagornov, N.N.; Lyakhov, P.A.; Valuev, G.V.; Chervyakov, N.I. (2020). "Application of the residue number system to reduce hardware costs of the convolutional neural network implementation". Mathematics and Computers in Simulation. Elsevier BV. 177: 232–243. doi: 10.1016/j.matcom.2020.04.031. ISSN 0378-4754. Convolutional neural networks are a promising tool for solving the problem of pattern recognition.
3. Zhang, Wei (1988). "Shift-invariant pattern recognition neural network and its optical architecture". Proceedings of Annual Conference of the Japan Society of Applied Physics.
4. COVID-19 pandemic – Wikipedia

(<https://en.wikipedia.org/wiki/COVID-19_pandemic>)

1. Lưu ý về đeo khẩu trang ngừa lây nhiễm COVID-19 - Bộ Y tế - Trang tin về dịch bệnh viêm đường hô hấp cấp COVID-19

(<https://ncov.moh.gov.vn/en/web/guest/-/luu-y-ve-eo-khau-trang-ngua-lay-nhiem-covid-19>)

1. H. A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," in IEEE Transactions on Pattern Analysis and Machine Intelligence.
2. Ševo and A. Avramović, "Convolutional Neural Network Based Automatic Object Detection on Aerial Images," in IEEE Geoscience and Remote Sensing Letters.

**Giáo viên hướng dẫn:** ThS. Lương Trần Ngọc Khiết.  
(Số điện thoại: 0387469465 - Email: [khietltn@hcmue.edu.vn](mailto:khietltn@hcmue.edu.vn) )

Nhóm sinh viên thực hiện:

* Lê Tấn Lộc

(Số điện thoại: 0785607050 - Email: [ah09program@gmail.com](mailto:ah09program@gmail.com))

* Huỳnh Thanh Phong

(Số điện thoại: 0931493972 - Email: [huynhthanhphong0931@gmail.com](mailto:huynhthanhphong0931@gmail.com))

* Phạm Duy Minh

(Số điện thoại: 0333454717 - Email: [duyminh081201@gmail.com](mailto:duyminh081201@gmail.com))

* Trần Thị Tứ Linh

(Số điện thoại: 0353544089 - Email: [tranthitulinh1305@gmail.com](mailto:tranthitulinh1305@gmail.com))