

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**NIÊN LUẬN CÔNG NGHỆ THÔNG TIN**

**ĐỀ TÀI:  
ÁP DỤNG MÁY HỌC VÀO PHÁT HIỆN NGƯỜI  
KHÔNG ĐEO KHẨU**

Sinh viên thực hiện:  
**PHAN THỊ ÁNH NGUYỆT**  
**MSSV: B1809269**

**Khóa: 44**

Cần Thơ, 10/2021

**BỘ GIÁO DỤC VÀ ĐÀO TẠO  
TRƯỜNG ĐẠI HỌC CẦN THƠ  
KHOA CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**NIÊN LUẬN CÔNG NGHỆ THÔNG TIN**

**ĐỀ TÀI:  
ÁP DỤNG MÁY HỌC VÀO PHÁT HIỆN NGƯỜI  
KHÔNG ĐEO KHẨU**

**Giáo viên hướng dẫn:  
TS. THÁI MINH TUẤN**

**Sinh viên thực hiện:  
PHAN THỊ ÁNH NGUYỆT  
MSSV: B1809269  
Khóa: 44**

**Cần Thơ, 10/2021**

## NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN



.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

Cần Thơ, ngày 10 tháng 12 năm 2021

**Giáo viên hướng dẫn**

**TS. Thái Minh Tuấn**

## LỜI CẢM ƠN



Sau một thời gian dài học tập và nghiên cứu, cuối cùng em đã hoàn thành đề tài niên luận này, đây là dịp tốt nhất để em có thể gửi lời cảm ơn đến mọi người.

Em xin gửi lời cảm ơn sâu sắc đến giáo viên hướng dẫn - Thầy Thái Minh Tuấn, người đã giảng dạy tận tình, chi tiết để em có đủ kiến thức và vận dụng vào bài niên luận này.

Xin gửi lời cảm ơn chân thành đến tất cả bạn bè đã giúp đỡ, giới thiệu những nguồn tài liệu quý giá để tôi có thể hoàn thành đề tài này.

Do chưa có nhiều kinh nghiệm làm đề tài cũng như những hạn chế về kiến thức, trong bài niên luận chắc chắn sẽ không tránh khỏi những thiếu sót. Rất mong nhận được sự nhận xét, ý kiến đóng góp, phê bình từ phía Thầy để bài niên luận được hoàn thiện hơn.

Em xin chân thành cảm ơn!

Cần Thơ, ngày 10 tháng 12 năm 2021

Sinh viên thực hiện

Phan Thị Ánh Nguyệt

# MỤC LỤC

MỤC LỤC .....	iii
TÓM TẮT .....	iv
PHẦN 1: TỔNG QUAN .....	1
I.    ĐẶT VẤN ĐỀ .....	1
II.   MỤC TIÊU ĐỀ TÀI.....	1
III.  BỐ CỤC NIÊN LUẬN.....	1
PHẦN 2: NỘI DUNG .....	3
CHƯƠNG 1: CƠ SỞ LÝ THUYẾT.....	3
1.1   MẠNG NƠON .....	3
1.1.1  Giới thiệu về mạng nơon .....	3
1.1.2  Kiến trúc chung của mạng nơon .....	4
1.2   MẠNG NƠON TÍCH CHẬP (CNN) .....	6
1.2.1  Giới thiệu về mạng nơon tích chập .....	6
1.2.2  Cấu trúc của mạng nơon tích chập .....	7
1.2.3  Xây dựng mạng nơon tích chập.....	8
CHƯƠNG 2: XÂY DỰNG HỆ THỐNG .....	12
2.1   ĐỀ XUẤT KIẾN TRÚC .....	12
2.2   XÂY DỰNG MÔ HÌNH .....	12
2.2.1  Tập dữ liệu .....	12
2.2.2  Tóm tắt quá trình xây dựng .....	12
2.2.3  Kết quả thực nghiệm.....	13
2.2.4  Đánh giá .....	13
PHẦN 3: KẾT LUẬN.....	14
I.    KẾT QUẢ ĐẠT ĐƯỢC .....	14
II.   HẠN CHẾ .....	14
III.  HƯỚNG PHÁT TRIỂN .....	14
TÀI LIỆU THAM KHẢO.....	15

## TÓM TẮT

Việc đeo khẩu trang nơi công cộng đã góp phần hạn chế được sự lây lan của dịch bệnh Covid 19 trên toàn cầu. Tuy nhiên, vẫn còn nhiều người dân chủ quan, thờ ơ không đeo khẩu trang nơi công cộng, đây là một trong những nguyên nhân gây ra sự lây lan dịch bệnh. Giám sát người dân thực hiện đúng chủ trương, chính sách của Nhà nước trong việc đeo khẩu trang nơi công cộng hoàn toàn có thể được thực hiện một cách tự động. Bằng việc ứng dụng công nghệ máy học và xử lý ảnh, đề tài đã xây dựng chương trình phát hiện người dân có đeo khẩu trang hay không trên video trực tiếp. Trong đó, em đã thực hiện thu thập dữ liệu để xây dựng mô hình đào tạo dựa trên mạng nơ-ron tích chập (CNN) với kiến trúc MobileNetV2. Chương trình triển khai trên ngôn ngữ Python và sử dụng một số thư viện mã nguồn mở như OpenCV, Tensorflow, Keras,... Có thể được tích hợp với các hệ thống nhúng để ứng dụng trong sân bay, nhà ga, văn phòng, trường học và những nơi công cộng để đảm bảo tuân thủ các nguyên tắc an toàn công cộng.

## PHẦN 1: TỔNG QUAN

### I. ĐẶT VẤN ĐỀ

Ngày 11/3/2020, Tổ chức Y tế Thế giới (WHO) đã ra tuyên bố gọi COVID-19 là "Đại dịch toàn cầu". Để ngăn chặn tình trạng lây lan nhanh chóng của đại dịch, bên cạnh khuyến nghị mà WHO đưa ra về việc đeo khẩu trang ở nơi đông người, Chính phủ Việt Nam cũng đã yêu cầu người dân phải đeo khẩu trang tại các điểm công cộng để hạn chế sự lây lan của dịch bệnh. Khẩu trang giúp hạn chế việc hít thở trực tiếp các giọt không khí có chứa virus và các tác nhân gây bệnh khác hoặc khi tiếp xúc với người mang mầm bệnh; việc đeo khẩu trang cũng giúp ngăn ngừa virus xâm nhập trực tiếp qua đường hít thở khi người đó hắt hơi, ho hoặc nói chuyện. Nhiều người đeo khẩu trang nhưng không đúng quy định, không có tác dụng phòng, chống dịch bệnh, tiềm ẩn rất nhiều nguy cơ lây nhiễm. Để giám sát người dân thực hiện theo đúng chỉ đạo của Chính phủ theo hình thức cũ là khá khó khăn và tốn kém vì thiếu nguồn nhân lực để giám sát.

Với sự phát triển nhanh chóng của học sâu (một chi của ngành học máy), đặc biệt là mô hình mạng nơron tích chập (CNN), thị giác máy tính đã đạt được những tiến bộ đáng kể trong những năm gần đây về nhận dạng và phát hiện đối tượng. CNN là nền tảng siêu quan trọng trong hầu hết các task trong lĩnh vực computer vision trong một thời gian dài. Tất nhiên, từ vanilla CNN, đã có rất nhiều (siêu nhiều) những cải tiến liên tiếp. Nhắc tới kiến trúc mạng nơron tích chập, ta có thể kể ra danh sách dài đằng đẵng: LeNet, AlexNet, VGG, GoogLeNet, ResNet... Mỗi kiến trúc CNN đều có thể mạnh của riêng nó, được điều chỉnh để phù hợp với các mục đích khác nhau. Khá nhiều kiến trúc CNN với mục đích gọn nhẹ có thể kể đến MobileNets, ShuffleNet, EffNet... MobileNets dựa trên một kiến trúc được sắp xếp hợp lý sử dụng các tập hợp có thể phân tách theo chiều sâu để xây dựng các mạng thần kinh sâu có trọng lượng nhẹ, để giảm số lượng tính toán, giảm số lượng params, đồng thời có thể thực hiện trích xuất đặc trưng một cách tách biệt trên các channel khác nhau. Từ thực trạng trên và áp dụng kiến thức sau khi nghiên cứu về mạng nơron tích chập, em đã chọn đề tài “ Phát hiện người không đeo khẩu trang” để thực hiện niên luận của mình, nhằm góp phần nhắc nhở, quản lý người đeo khẩu trang, cùng chung tay nâng cao ý thức cộng đồng, đẩy lùi đại dịch COVID-19.

### II. MỤC TIÊU ĐỀ TÀI

Xây dựng hệ thống phát hiện người không đeo khẩu trang với độ chính xác cao.

Nghiên cứu và tìm hiểu về mạng nơron tích chập CNN.

### III. BỐ CỤC NIÊN LUẬN

#### Phần 1: Tổng quan

Tổng quan về đề tài

#### Phần 2: Nội dung

**Chương 1:** Cơ sở lý thuyết

**Chương 2:** Xây dựng hệ thống

**Chương 3:** Kết quả thực nghiệm

**Phần 3: Kết luận**

Trình bày kết quả đạt được và hướng phát triển hệ thống.



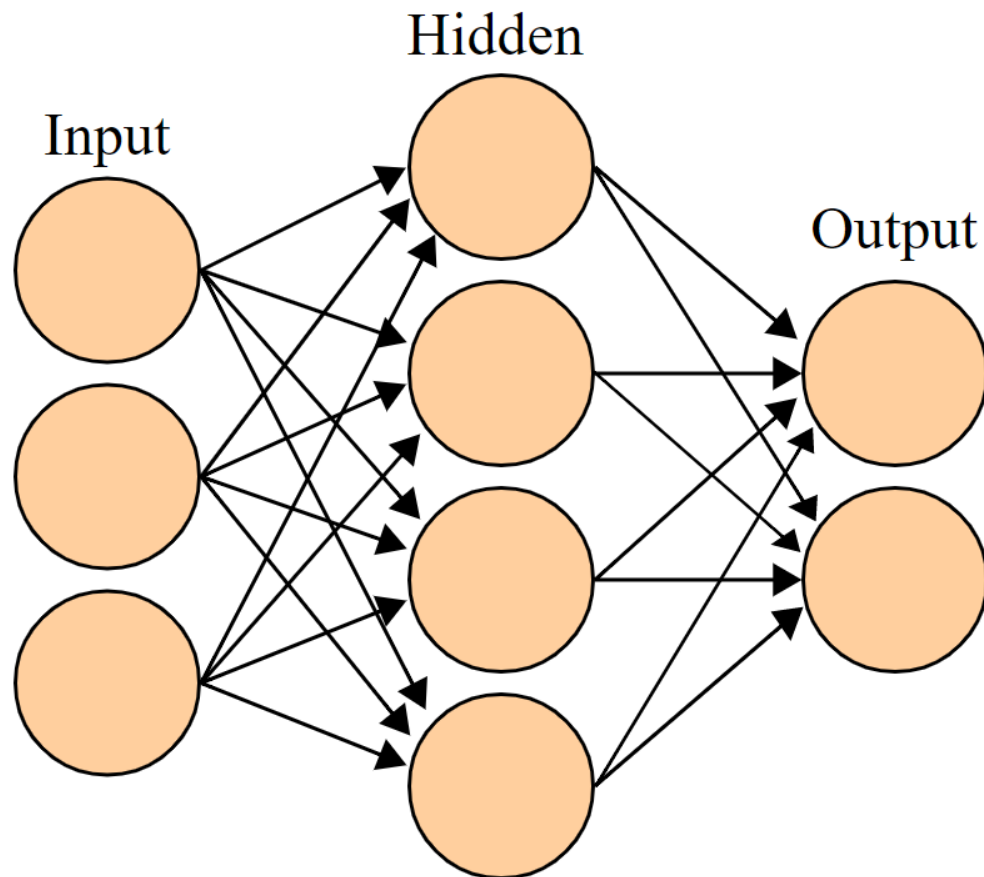
## PHẦN 2: NỘI DUNG

### CHƯƠNG 1: CƠ SỞ LÝ THUYẾT

#### 1.1 MẠNG NƠRON

##### 1.1.1 Giới thiệu về mạng nơron

Mạng neural nhân tạo hay thường gọi ngắn gọn là mạng neural (tiếng Anh là artificial neural network - ANN hay neural network) là một mô hình toán học hay mô hình tính toán được xây dựng dựa trên các mạng neural sinh học. Nó gồm có một nhóm các neural nhân tạo (nút) nối với nhau, và xử lý thông tin bằng cách truyền theo các kết nối và tính giá trị mới tại các nút (cách tiếp cận connectionism đối với tính toán). Trong nhiều trường hợp, mạng neural nhân tạo là một hệ thống thích ứng (adaptive system) tự thay đổi cấu trúc của mình dựa trên các thông tin bên ngoài hay bên trong chảy qua mạng trong quá trình học.

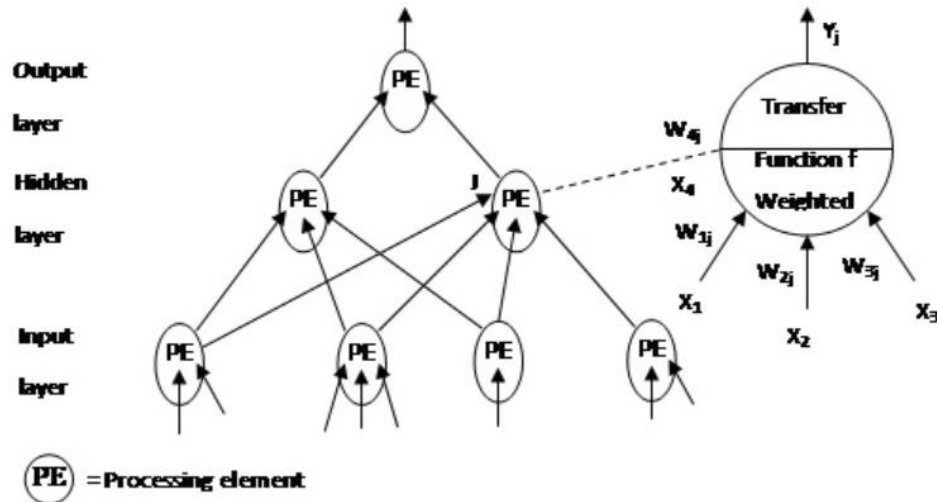


*Hình 1: Minh họa mạng nơron*

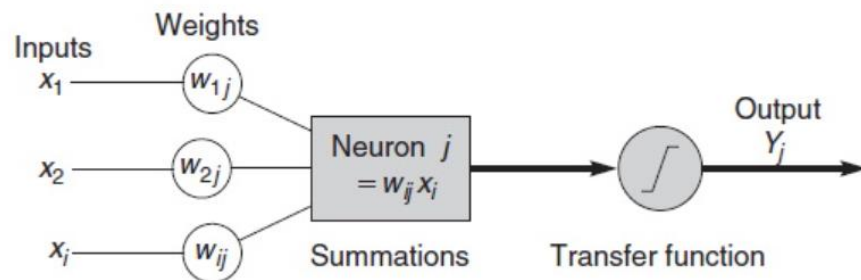
Trong thực tế sử dụng, nhiều mạng neural là các công cụ mô hình hóa dữ liệu thống kê phi tuyến. Chúng có thể được dùng để mô hình hóa.

### 1.1.2 Kiến trúc chung của mạng nơron

Kiến trúc chung của một mạng nơron nhân tạo (ANN) gồm 3 thành phần đó là: Input Layer, Hidden Layer và Output Layer.



- Lớp ẩn (Hidden Layer) gồm các Nơron nhận dữ liệu input từ các Nơron ở lớp (Layer) trước đó và chuyển đổi các input này cho các lớp xử lý tiếp theo. Trong một ANN có thể có nhiều lớp ẩn.
- Các Processing Elements (PE) của ANN gọi là Nơron, mỗi Nơron nhận các dữ liệu vào (Inputs) xử lý chúng và cho ra một kết quả (Output) duy nhất. Kết quả xử lý của một Nơron có thể làm Input cho các Nơron khác.
- Quá trình xử lý thông tin của một ANN:



- + Inputs (dữ liệu vào): Mỗi Input tương ứng với 1 thuộc tính (attribute) của dữ liệu (patterns).
- + Output (kết quả): Kết quả của một ANN là một giải pháp cho một vấn đề.
- + Connection Weights (Trọng số liên kết): Đây là thành phần rất quan trọng của một ANN, nó thể hiện mức độ quan trọng (độ

mạnh) của dữ liệu đầu vào đối với quá trình xử lý thông tin (quá trình chuyển đổi dữ liệu từ Layer này sang layer khác). Quá trình học (Learning Processing) của ANN thực ra là quá trình điều chỉnh các trọng số (Weight) của các input data để có được kết quả mong muốn.

- + **Summation Function (Hàm tổng):** Tính tổng trọng số của tất cả các input được đưa vào mỗi Noron (phần tử xử lý PE). Hàm tổng của một Noron đối với n input được tính theo công thức sau:

$$Y = \sum_{i=1}^n X_i W_i$$

- + **Transfer Function (Hàm chuyển đổi):** Hàm tổng (Summation Function) của một Noron cho biết khả năng kích hoạt (Activation) của Noron đó còn gọi là kích hoạt bên trong (internal activation). Các Noron này có thể sinh ra một output hoặc không trong ANN (nói cách khác rằng có thể output của 1 Noron có thể được chuyển đến layer tiếp trong mạng Noron hoặc không). Mối quan hệ giữa Internal Activation và kết quả (output) được thể hiện bằng hàm chuyển đổi (Transfer Function).

Việc lựa chọn Transfer Function có tác động lớn đến kết quả của ANN. Hàm chuyển đổi phi tuyến được sử dụng phổ biến trong ANN là sigmoid (logical activation) function.

$$YT = 1 / (1 + e^{-Y})$$

Trong đó:

$Y_T$ : Hàm chuyển đổi

$Y$ : Hàm tổng

Kết quả của Sigmoid Function thuộc khoảng [0,1] nên còn gọi là hàm chuẩn hóa (Normalized Function).

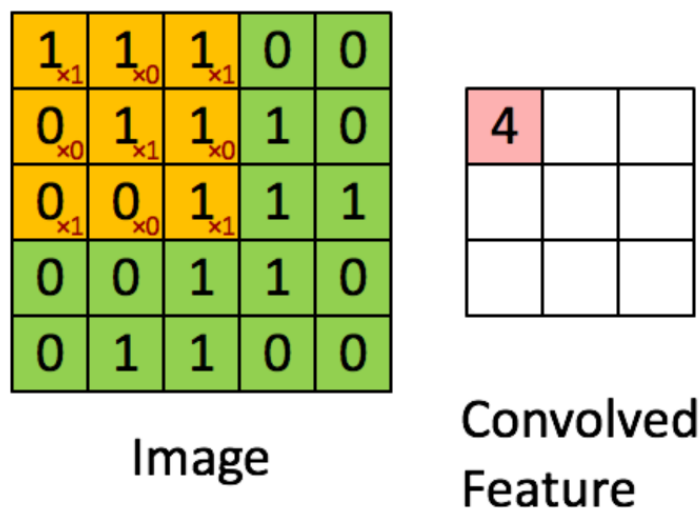
Kết quả xử lý tại các Noron (Output) đôi khi rất lớn, vì vậy transfer function được sử dụng để xử lý output này trước khi chuyển đến layer tiếp theo. Đôi khi thay vì sử dụng Transfer Function người ta sử dụng giá trị ngưỡng (Threshold value) để kiểm soát các output của các Noron tại một layer nào đó trước khi chuyển các output này đến các Layer tiếp theo. Nếu output của một noron nào đó nhỏ hơn giá trị ngưỡng thì nó sẽ không được chuyển đến Layer tiếp theo.

## 1.2 MẠNG NƠ-RO-N TÍCH CHẬP (CNN)

### 1.2.1 Giới thiệu về mạng nơ-ron tích chập

Convolutional Neural Network (CNNs – 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 với độ chính xác cao như hiện nay. CNN được sử dụng nhiều trong các bài toán nhận dạng các object trong ảnh.

Convolutional là một cửa sổ trượt (Sliding Windows) trên một ma trận như mô tả hình dưới:



**Hình 2:** Minh họa tích chập

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

Trong hình ảnh ví dụ trên, ma trận bên trái là một hình ảnh trắng đen được số hóa. Ma trận có kích thước 5×5 và mỗi điểm ảnh có giá trị 1 hoặc 0 là giao điểm của dòng và cột.

Convolution hay tích chập là nhân từng phần tử trong ma trận 3. Sliding Window hay còn gọi là kernel, filter hoặc feature detect là một ma trận có kích thước nhỏ như trong ví dụ trên là 3×3.

Convolution hay tích chập là nhân từng phần tử bên trong ma trận 3×3 với ma trận bên trái. Kết quả được một ma trận gọi là Convolved feature được sinh ra từ việc nhân ma trận Filter với ma trận ảnh 5×5 bên trái.

### 1.2.2 Cấu trúc của mạng nơron tích chập

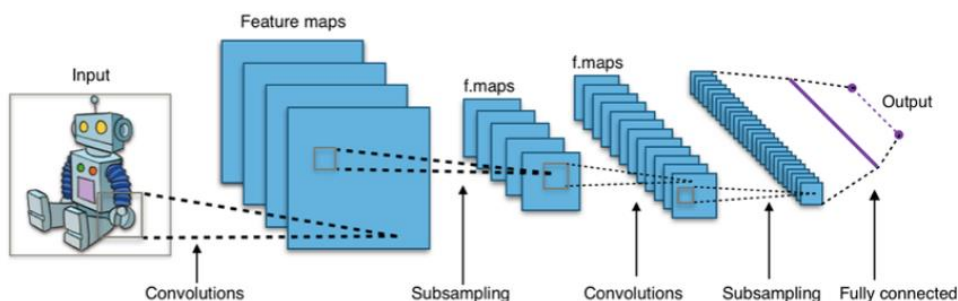
Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation như ReLU và tanh để kích hoạt các trọng số trong các node. Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.

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. Trong mô hình mạng truyền ngược (feedforward neural network) thì mỗi neural đầu vào (input node) cho mỗi neural đầu ra trong các lớp tiếp theo.

Mô hình này gọi là mạng kết nối đầy đủ (fully connected layer) hay mạng toàn vẹn (affine layer). Còn trong mô hình CNNs thì ngược lại. Các layer liên kết được với nhau thông qua cơ chế convolution.

Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. 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:** Mô hình mạng nơron tích chập

Trong quá trình huấn luyện mạng (training) CNN tự động học các giá trị qua các lớp filter dựa vào cách thức mà bạn thực hiện. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.

Trong mô hình CNN có 2 khía cạnh cần quan tâm là tính bất biến (Location Invariance) và tính kết hợp (Compositionality). Với cùng một đối tượng, nếu đối tượng này được chiếu theo các góc độ khác nhau (translation, rotation, scaling) thì độ chính xác của thuật toán sẽ bị ảnh hưởng đáng kể.

Pooling layer sẽ cho bạn tính bất biến đối với phép dịch chuyển (translation), phép quay (rotation) và phép co giãn (scaling). Tính kết hợp cục bộ cho ta các cấp độ biểu diễn thông tin từ mức độ thấp đến mức độ cao và trừu tượng hơn thông qua convolution từ các filter.

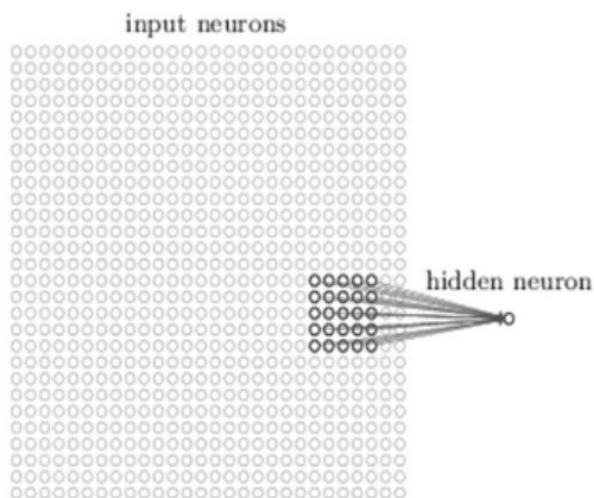
### 1.2.3 Xây dựng mạng neuron tích chập

Mạng CNN sử dụng 3 ý tưởng cơ bản:

- **Các trường tiếp nhận cục bộ (local receptive field)**

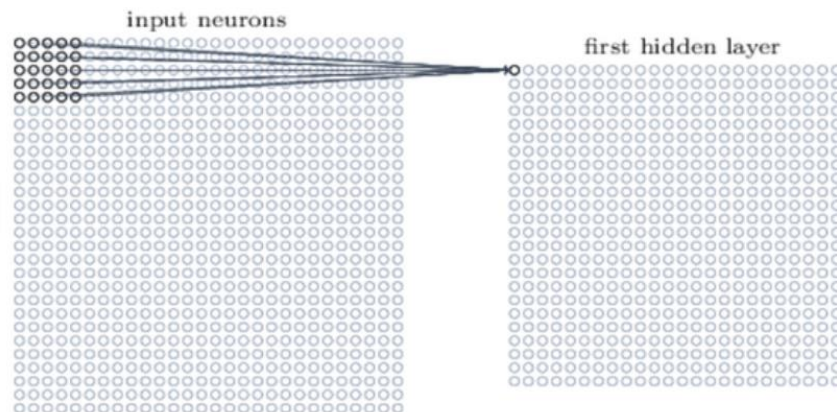
Đầu vào của mạng CNN là một ảnh. Ví dụ như ảnh có kích thước  $28 \times 28$  thì tương ứng đầu vào là một ma trận có  $28 \times 28$  và giá trị mỗi điểm ảnh là một ô trong ma trận. Trong mô hình mạng ANN truyền thống thì chúng ta sẽ kết nối các neuron đầu vào vào tầng ảnh.

Tuy nhiên trong CNN chúng ta không làm như vậy mà chúng ta chỉ kết nối trong một vùng nhỏ của các neuron đầu vào như một filter có kích thước  $5 \times 5$  tương ứng  $(28 - 5 + 1) = 24$  điểm ảnh đầu vào. Mỗi một kết nối sẽ học một trọng số và mỗi neuron ẩn sẽ học một bias. Mỗi một vùng  $5 \times 5$  đây gọi là một trường tiếp nhận cục bộ.

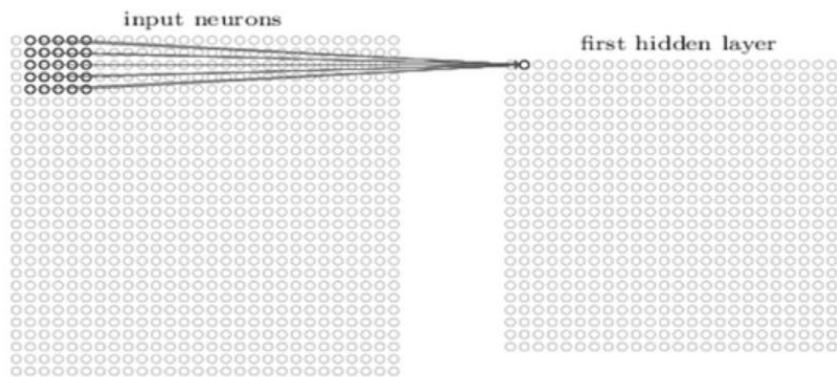


Một cách tổng quan, ta có thể tóm tắt các bước tạo ra 1 hidden layer bằng các cách sau:

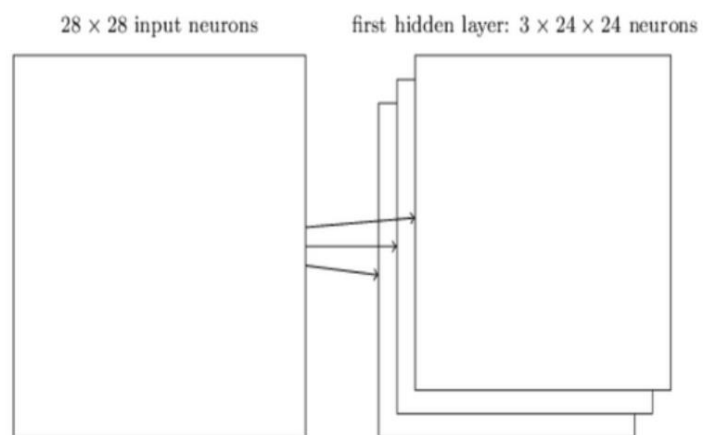
### 1. Tạo ra neuron ẩn đầu tiên trong lớp ẩn 1



### 2. Dịch filter qua bên phải một cột sẽ tạo được neuron ẩn thứ 2.



với bài toán nhận dạng ảnh người ta thường gọi ma trận lớp đầu vào là feature map, trọng số xác định các đặc trưng là shared weight và độ lệch xác định một feature map là shared bias. Như vậy đơn giản nhất là qua các bước trên chúng ta chỉ có 1 feature map. Tuy nhiên trong nhận dạng ảnh chúng ta cần nhiều hơn một feature map.





Như vậy, local receptive field thích hợp cho việc phân tách dữ liệu ảnh, giúp chọn ra những vùng ảnh có giá trị nhất cho việc đánh giá phân lớp.

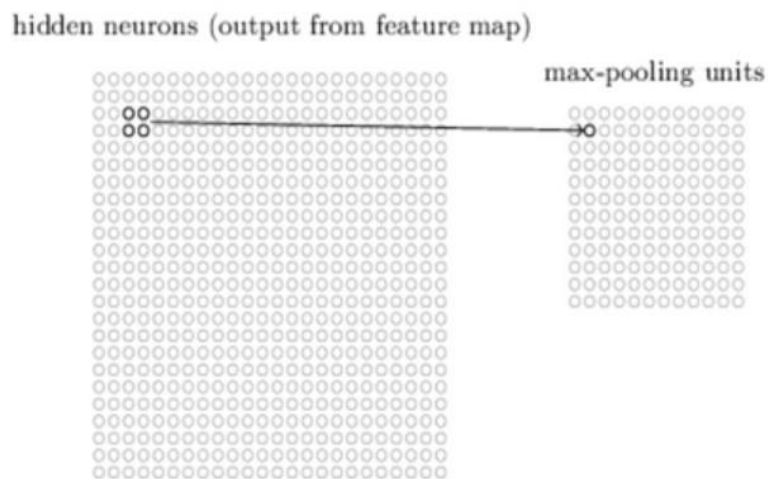
– **Trọng số chia sẻ (shared weights and bias)**

Đầu tiên, các trọng số cho mỗi filter (kernel) phải giống nhau. Tất cả các nơ-ron trong lớp ẩn đầu sẽ phát hiện chính xác feature tương tự chỉ ở các vị trí khác nhau trong hình ảnh đầu vào. Chúng ta gọi việc map từ input layer sang hidden layer là một feature map.

Tóm lại, một convolutional layer bao gồm các feature map khác nhau. Mỗi một feature map giúp detect một vài feature trong bức ảnh. Lợi ích lớn nhất của trọng số chia sẻ là giảm tối đa số lượng tham số trong mạng CNN.

– **Lớp tổng hợp (pooling layer).**

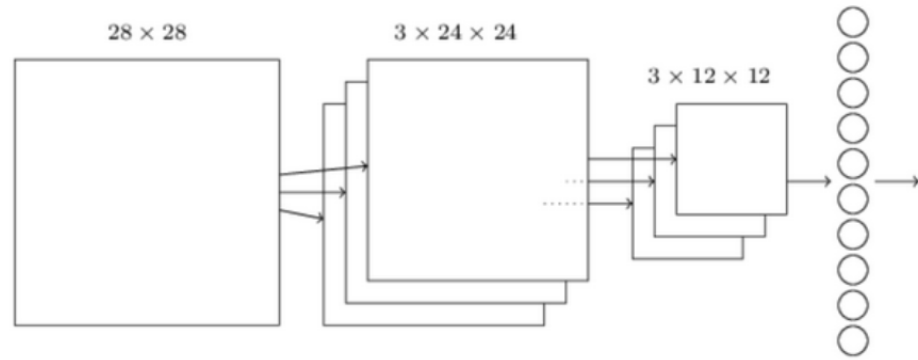
Lớp pooling thường được sử dụng ngay sau lớp convulational để đơn giản hóa thông tin đầu ra để giảm bớt số lượng neuron. Thủ tục pooling phổ biến là max-pooling, thủ tục này chọn giá trị lớn nhất trong vùng đầu vào  $2 \times 2$ .



Như vậy qua lớp Max Pooling thì số lượng neuron giảm đi phân nửa. Trong một mạng CNN có nhiều Feature Map nên mỗi Feature Map chúng ta sẽ cho mỗi Max Pooling khác nhau. Chúng ta có thể thấy rằng Max Pooling là cách hỏi xem trong các đặc trưng này thì đặc trưng nào là đặc trưng nhất. Ngoài Max Pooling còn có L2 Pooling.

Cuối cùng ta đặt tất cả các lớp lại với nhau thành một CNN với đầu ra gồm các neuron với số lượng tùy bài toán.





Hai lớp cuối cùng của các kết nối trong mạng là một lớp đầy đủ kết nối (fully connected layer). Lớp này nối mọi nơon từ lớp max pooled tới mọi nơon của tầng ra.

---

## CHƯƠNG 2: XÂY DỰNG HỆ THỐNG

### 2.1 ĐỀ XUẤT KIẾN TRÚC

Trong đề tài này, em đề xuất xây dựng mô hình đào tạo dựa trên mạng nơron tích chập (CNN) với kiến trúc MobileNet.

MobileNet sử dụng cách tính chập tích mang tên DSC (Depthwise Separable Convolution) nhằm giảm kích thước mô hình và giảm độ phức tạp. Do đó, MobileNet thường được sử dụng cho Computer Vision ứng dụng trên các thiết bị nhỏ gọn như điện thoại thông minh hoặc thiết bị nhúng.

Depthwise convolution là một loại tích chập trong đó chúng ta áp dụng một bộ lọc tích chập duy nhất cho mỗi kênh đầu vào. Trong phép tích chập 2D thông thường được thực hiện trên nhiều kênh đầu vào, bộ lọc cũng sâu như đầu vào và cho phép chúng ta tự do trộn các kênh để tạo ra từng phần tử trong đầu ra. Ngược lại, sự biến đổi theo chiều sâu giữ cho mỗi kênh riêng biệt.

### 2.2 XÂY DỰNG MÔ HÌNH

#### 2.2.1 Tập dữ liệu

Tập dữ liệu bao gồm 3833 ảnh, trong đó có 1915 ảnh cho các trường hợp đeo trang và 1918 ảnh cho trường hợp không đeo khẩu trang. Bộ ảnh dùng để test chiếm 20% khoảng 767 ảnh và bộ ảnh để huấn luyện gồm 3066 ảnh. Tập dữ liệu được tham khảo tại <https://github.com/balajisrinivas/Face-Mask-Detection>.

Tập dữ liệu gồm những hình ảnh có định dạng “.jpg”, “.jpeg”, “.png” và nhãn của các hình ảnh là “with\_mask”, “without\_mask”.

#### 2.2.2 Tóm tắt quá trình xây dựng

Xây dựng cấu trúc chương trình gồm 3 bước sau:

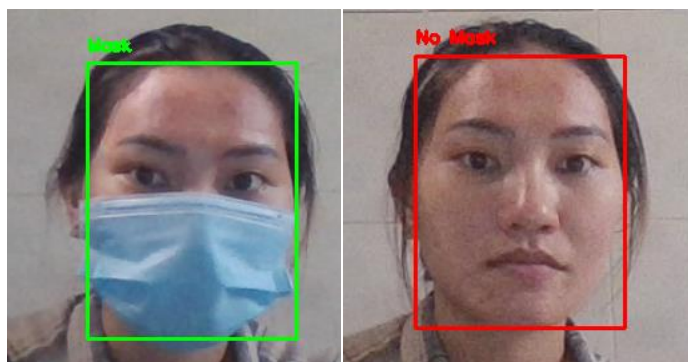
- **Bước 1:** thu thập dữ liệu chương trình bằng Python (ngôn ngữ lập trình bậc cao), sử dụng thư viện phần mềm mã nguồn mở OpenCV để phát hiện khuôn mặt người. Dữ liệu sau khi thu thập dưới dạng file ảnh (JPG, JPEG, PNG) sẽ được lưu trữ ở hai thư mục riêng biệt gồm: một thư mục chứa 1915 bức ảnh mô tả khuôn mặt đeo khẩu trang, thư mục còn lại chứa 1918 bức ảnh mô tả khuôn mặt không đeo khẩu trang.
- **Bước 2:** sử dụng nguồn dữ liệu đã thu thập được ở bước 1 để phân tích dựa trên mô hình CNN. Ở giai đoạn này, xử lý tiền dữ liệu nhằm đưa tất cả các ảnh về cùng kích thước, sau đó các ảnh này sẽ được chuyển đổi để phục vụ cho quá trình xử lý ảnh ở bước sau. Dựa vào mô hình CNN, các nơron tích chập được thiết kế đặc biệt để xử lý các

phần tử quan trọng nhất trên bức ảnh nhằm đưa ra kết quả dữ liệu chính xác.

- **Bước 3:** phát hiện người đeo khẩu trang hay không. Bước này sẽ tiến hành phân tích so sánh dữ liệu được trích xuất từ camera (sau khi đã được xử lý dữ liệu đầu vào) với kết quả dữ liệu đã được phân tích để cảnh báo. Dựa vào kết quả thu được từ bước 2, dữ liệu sẽ hiển thị lên màn hình kết quả người dân có đeo khẩu trang hay không.

### 2.2.3 Kết quả thực nghiệm

Nhận diện từ camera thực tế:



*Hình 4: Kết quả nhận diện từ camera*

### 2.2.4 Đánh giá

Các chỉ số đánh giá thu được từ tập test 767 ảnh:

	precision	recall	f1-score	support
with_mask	0.99	0.99	0.99	383
without_mask	0.99	0.99	0.99	384
accuracy			<b>0.99</b>	<b>767</b>

Trong đó, mô hình dự đoán được 383 ảnh cho trường hợp đeo khẩu trang và 384 ảnh cho trường hợp không đeo khẩu trang.

Độ chính xác cho từng phân lớp là 0.99. Độ chính xác trung bình là 99%.

---

## **PHẦN 3: KẾT LUẬN**

### **I. KẾT QUẢ ĐẠT ĐƯỢC**

Đề tài đã xây dựng được một hệ thống phát hiện người không đeo khẩu trang trên video trực tiếp với độ chính xác cao, cài đặt tương đối nhanh.

Tìm hiểu nâng cao kiến thức về mạng nơron và mạng nơron tích chập, củng cố nền tảng trong lĩnh vực học sâu.

### **II. HẠN CHẾ**

Song song những kết quả đạt được, vẫn còn tồn tại những hạn chế: Mô hình chỉ phát hiện được người có đeo khẩu trang hay không, chưa phát hiện được người đeo khẩu trang không đúng cách.

### **III. HƯỚNG PHÁT TRIỂN**

Tiếp tục phát triển hệ thống hoàn thiện, phát hiện được người đeo khẩu trang không đúng cách. Xây dựng bổ sung thêm hệ thống nhắc nhở để có thể áp dụng hệ thống vào thực tế.

---

## TÀI LIỆU THAM KHẢO

### Tài liệu:

- [1] **Huỳnh Triệu Vĩ**. Ước lượng số lượng của giống sử dụng phương pháp CNN-Based Crowd Counting, Trường Đại học Cần Thơ, 2020.
- [2] <http://vjst.vn/vn/tin-tuc/5323/phat-hien-nguoi-deo-khau-trang-trong-thoi-gian-thuc.aspx>
- [3] <https://topdev.vn/blog/thuat-toan-cnn-convolutional-neural-network/>
- [4] [http://www.nawapi.gov.vn/index.php?option=com\\_content&view=article&id=3238%3Agii-thiu-tng-quan-v-mng-nron-nhan-to-artificial-neural-network-ann&catid=70%3Anhim-v-chuyen-mon-ang-thc-hin&Itemid=135&lang=vi](http://www.nawapi.gov.vn/index.php?option=com_content&view=article&id=3238%3Agii-thiu-tng-quan-v-mng-nron-nhan-to-artificial-neural-network-ann&catid=70%3Anhim-v-chuyen-mon-ang-thc-hin&Itemid=135&lang=vi)
- [5] <https://github.com/balajisrinivas/Face-Mask-Detection>