

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

-----



**Bùi Việt Anh**

**MÔ HÌNH HỌC SÂU VÀ ỨNG DỤNG CHO BÀI TOÁN NHẬN  
DẠNG HÌNH ẢNH TRONG THƯƠNG MẠI ĐIỆN TỬ**

CHUYÊN NGÀNH : HỆ THỐNG THÔNG TIN  
MÃ SỐ : 60.48.01.04

**TÓM TẮT LUẬN VĂN THẠC SĨ KỸ THUẬT**

HÀ NỘI - 2022

Luận văn được hoàn thành tại:

**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

Người hướng dẫn khoa học: PGS.TS. Nguyễn Văn Thủy

Phản biện 1: PGS.TS. Phạm Thanh Giang

Phản biện 2: TS. Nguyễn Vĩnh An

Luận văn sẽ được bảo vệ trước Hội đồng chấm luận văn thạc sĩ tại Học viện Công nghệ Bưu chính Viễn thông

Vào lúc: 10 giờ 45 phút, ngày 2 tháng 7 năm 2022

Có thể tìm hiểu luận văn tại:

- Thư viện của Học viện Công nghệ Bưu chính Viễn thông.

## MỞ ĐẦU

### 1. Tính cấp thiết của đề tài

Cùng với sự phát triển bùng nổ của internet, các kênh tìm hiểu thông tin, giải trí, thương mại điện tử... cùng phát triển nhanh chóng và mạnh mẽ. Giờ đây, gần như chúng ta có thể tìm kiếm được mọi thứ trên internet, từ tài liệu, sách, truyện, phim, quần, áo, giày,...

Sử dụng internet chính là muốn mua bán, giảm bớt khâu tìm kiếm bằng từ khóa truyền thống vốn kém hiệu quả, tích hợp các trợ lý ảo hỗ trợ người dùng sử dụng đa ngôn ngữ trên toàn cầu, nâng cao trải nghiệm mua sắm của khách hàng cho đến nâng cao hiệu quả của các công cụ phân tích dự báo. Nhận dạng hình ảnh sản phẩm có khả năng giúp các thương hiệu phát hiện bất kỳ người nào chia sẻ nội dung hình ảnh liên quan đến sản phẩm của họ trên các sàn thương mại điện tử dù họ không nhắc đến tên các thương hiệu này, đồng thời công nghệ này còn cho phép tính cá nhân hóa trong trải nghiệm mua hàng. Vì vậy, hệ thống nhận dạng hình ảnh trong thương mại điện tử giải quyết các vấn đề đó.

### 2. Tình hình nghiên cứu và tính mới của đề tài

Trong thời đại công nghệ như hiện nay việc ứng dụng nhận dạng hình ảnh là rất cần thiết đối với người sử dụng các sàn thương mại điện tử. Từ trước đến nay, cùng với sự tiến bộ của khoa học công nghệ, đặc biệt là khả năng xử lý nhanh của phần mềm tự động, rất nhiều nghiên cứu đã tiến hành nhận dạng hình ảnh tự động trên nhiều môi trường khác nhau - “Sách trắng thương mại điện tử 2018”.

Đề tài tìm hiểu và ứng dụng nhận dạng hình ảnh cũng như cách triển khai công cụ tìm kiếm hình ảnh phần mềm tự động để giảm nguồn nhân lực và đảm bảo chất lượng phần hơn với công việc tìm kiếm bằng tay.

### 3. Mục tiêu nghiên cứu

Mục tiêu chính của đề tài là mô hình học sâu và ứng dụng cho nhận dạng hình ảnh trong thương mại điện tử để đạt được tốc độ tìm kiếm nhanh và chuẩn xác nhất để cho người dùng không mất nhiều thời gian tìm kiếm sản phẩm.

- Nghiên cứu về các hệ thống nhận dạng hình ảnh.
- Thử nghiệm, đánh giá độ hiệu quả của các thuật toán.
- Xây dựng hệ thống nhận dạng hình ảnh tự động giới thiệu sản phẩm.

### 4. Phương pháp nghiên cứu

#### 4.1. Nghiên cứu lý thuyết

Đọc và phân tích tài liệu về các phương pháp, thuật toán đã từng được sử dụng để xây dựng hệ thống nhận dạng hình ảnh.

#### **4.2. Nghiên cứu thực nghiệm**

- + Thử nghiệm và đánh giá độ hiệu quả của các thuật toán.
- + Xây dựng hệ thống nhận dạng hình ảnh trên một số sàn thương mại điện tử.

### **5. Bố cục của luận văn**

#### **Chương 1: Tổng quan về thương mại điện tử Việt Nam và thế giới**

Trình bày tổng quan về thương mại điện tử trên thế giới và ở Việt Nam. Làm rõ các thực trạng của thương mại điện tử ở nước ta và trên thế giới. Trình bày về hạ tầng thương mại điện tử ở nước ta. Đồng thời cũng nêu khái quát ứng dụng mô hình học sâu vào bài toán nhận dạng hình ảnh trong thương mại điện tử.

#### **Chương 3: Giới thiệu và nghiên cứu mạng CNN cho bài toán nhận dạng hình ảnh**

Khái quát về bài toán nhận dạng hình ảnh. Nghiên cứu về học máy, học sâu và về mạng nơ ron tích chập CNN và thành phần cụ thể của mạng này. Trình bày kiến trúc một số mạng, học chuyển giao. Nêu cụ thể về kiến trúc mạng được sử dụng phổ biến và rộng rãi hiện nay là ResNet.

#### **Chương 3: Thực nghiệm, đánh giá và xây dựng hệ thống nhận dạng hình ảnh trong thương mại điện tử**

Tiến hành thực nghiệm mô hình ResNet50 với tập dữ liệu hình ảnh các sản phẩm thương mại điện tử. Xây dựng trang web nhận dạng hình ảnh sản phẩm thương mại điện tử sử dụng mô hình học sâu ResNet50

## **CHƯƠNG I: TỔNG QUAN VỀ THƯƠNG MẠI ĐIỆN TỬ VIỆT NAM VÀ THẾ GIỚI.**

### **1.1 Giới thiệu về thương mại điện tử**

#### **1.1.1 Lịch sử phát triển**

Sự khởi đầu của thương mại điện tử có thể được bắt nguồn từ những năm 1960, khi các doanh nghiệp bắt đầu sử dụng EDI để chia sẻ tài liệu kinh doanh với các công ty khác.

Sau khi số lượng người dùng cá nhân chia sẻ tài liệu điện tử với nhau tăng lên vào những năm 1980, sự nổi lên của eBay và Amazon trong những năm 1990 đã tạo ra một cuộc cách mạng trong ngành thương mại điện tử. Giờ đây, người tiêu dùng có thể mua vô số mặt hàng trực tuyến, từ các cửa hàng bán hàng điện tử, cửa hàng truyền thống điển hình có khả năng thương mại điện tử. Hiện nay, hầu hết tất cả các công ty bán lẻ đang tích hợp các phương thức kinh doanh trực tuyến vào mô hình kinh doanh của họ.

#### **1.1.2 Khái niệm**

Khái niệm thương mại điện tử của WTO: “Thương mại điện tử (hay thương mại trực tuyến) bao gồm việc sản xuất, quảng cáo, bán hàng và phân phối sản phẩm được mua bán và thanh toán trên mạng Internet, nhưng được giao nhận một cách hữu hình, cả các sản phẩm giao nhận cũng như những thông tin số hoá thông qua mạng Internet”.

#### **1.1.3 Cách thức hoạt động**

Khi đơn đặt hàng được đặt, trình duyệt web của khách hàng sẽ giao tiếp qua lại với máy chủ lưu trữ trang web cửa hàng trực tuyến. Sau khi đơn hàng được xác thực, người quản lý đơn hàng sẽ thông báo đến máy chủ web của cửa hàng, sau đó sẽ hiển thị thông báo cho khách hàng rằng đơn hàng của họ đã được xử lý thành công. Sau đó, người quản lý đơn hàng sẽ gửi dữ liệu đơn hàng đến kho hàng hoặc bộ phận thực hiện để sản phẩm hoặc dịch vụ được gửi thành công cho khách hàng.

#### **1.1.4 Các hình thức**

Thị trường thương mại điện tử cũng được phân chia thành các hình thức khác nhau phụ thuộc vào từng đối tượng tham gia cụ thể. Hiện nay, có 6 loại hình thương mại điện tử cơ bản như sau: Doanh nghiệp với Doanh nghiệp (B2B), Doanh nghiệp với Khách hàng (B2C), Khách hàng với Khách hàng (C2C), Khách hàng với Doanh nghiệp (C2B), Doanh nghiệp với chính phủ (B2A), Khách hàng với Chính phủ (C2A)

#### **1.1.5 Đặc trưng**

Cụ thể Thương mại điện tử có các đặc trưng như sau: phổ biến, phạm vi tiếp cận toàn cầu, tiêu chuẩn chung, tính tương tác, mật độ thông tin, sự phong phú, cá nhân hóa

#### **1.1.6 Ưu và nhược điểm của thương mại điện tử**

- ❖ **Ưu điểm:** khả dụng, tốc độ truy cập, tính khả dụng rộng rãi, khả năng tiếp cận dễ dàng, tiếp cận quốc tế, chi phí thấp hơn, cá nhân hóa và đề xuất sản phẩm.

- ❖ **Nhược điểm:** dịch vụ khách hàng hạn chế, không thể chạm hoặc nhìn, thời gian chờ đợi.

### **1.1.7 Lợi ích thương mại điện tử đến doanh nghiệp**

Những lợi ích lớn như sau: rào cản nhập cảnh thấp, tăng thị phần tiềm năng, quảng cáo với chi phí thấp, lợi ích chiến lược, phạm vi tiếp cận toàn cầu.

## **1.2 Thực trạng thương mại điện tử trên thế giới và Việt Nam**

### **1.2.1 Thương mại điện tử toàn cầu.**

Năm 2007, thương mại điện tử chiếm 5,1% tổng mức bán lẻ; năm 2019, thương mại điện tử chiếm 16,0%. Sự phát triển vượt bậc này dẫn đến một xu thế hiện nay đó là thương mại điện tử toàn cầu

- **Thị trường thương mại điện tử toàn cầu**

Thị trường thương mại điện tử toàn cầu dự kiến đạt tổng trị giá 5,55 nghìn tỷ đô la vào năm 2022. Con số đó được ước tính sẽ tăng trong vài năm tới, cho thấy thương mại điện tử không biên giới đang trở thành một lựa chọn có lợi cho các nhà bán lẻ trực tuyến.

- **Tăng trưởng doanh số thương mại điện tử toàn cầu**

Vào năm 2023, với các trang web thương mại điện tử chiếm 22,3% tổng doanh số bán lẻ. Mặc dù bán lẻ đã có một năm khó khăn vào năm 2020, nhưng mọi thị trường quốc gia được eMarketer bao phủ đều chứng kiến mức tăng trưởng thương mại điện tử hai con số.

### **1.2.2 Thực trạng phát triển thương mại điện tử tại Việt Nam**

- **Nền tảng thương mại điện tử trong nước và quốc tế**

Thị trường thương mại điện tử của Việt Nam, trị giá 208,962 tỷ đồng (9 tỷ đô la Mỹ) vào năm 2019, kém 1 so với các thị trường phát triển hơn trong các báo cáo. Tuy nhiên, nó tự hào có một trong những dự đoán tăng trưởng cao hơn trong nghiên cứu, khi một quốc gia gồm những người tiêu dùng trẻ tuổi ngày càng chuyển sang các phương pháp kỹ thuật số để mua sắm.

- **Nhu cầu mua sắm trực tuyến của người Việt**

Mua sắm hàng tạp hóa trực tuyến đã có mức tăng trưởng sử dụng cao nhất so với các cửa hàng bán lẻ truyền thống. Thói quen mới và nhanh chóng được áp dụng này được kỳ vọng sẽ duy trì ngay cả sau đại dịch, cho thấy lĩnh vực thương mại điện tử đang tiếp tục mở rộng trong nước.

- **Triển vọng thương mại điện tử: tiền mặt so với thanh toán kỹ thuật số**

Trong số tất cả các lựa chọn thanh toán không dùng tiền mặt, ví điện tử đã và đang đạt được sức hút trong vài năm gần đây, theo đề xuất của số lượng người dùng và nhà cung cấp ngày càng tăng, cũng như ý định sử dụng trong tương lai cao của người Việt Nam.

## **1.3 Hạ tầng thương mại điện tử ở Việt Nam hiện nay**

### **1.3.1 Cơ sở hạ tầng pháp lý**

Vào tháng 9 năm 2020, Bộ Công Thương đã ban hành dự thảo thứ hai của nghị định sửa đổi Nghị định 52. Cụ thể, dự thảo thứ hai quy định rằng các đơn vị nước ngoài thiết lập trang web với miền Việt Nam hoặc thiết lập trang web thương mại điện tử có giao dịch / khách truy cập / đặt hàng từ Việt Nam, phải:

- (a) Đăng ký / thông báo các hoạt động thương mại điện tử của họ theo quy định của pháp luật Việt Nam,
- (b) Đảm bảo tuân thủ các quy định về bảo vệ quyền lợi người tiêu dùng và chịu trách nhiệm về chất lượng sản phẩm / hàng hóa do văn phòng đại diện hoặc đại diện được ủy quyền của mình phân phối qua website
- (c) Báo cáo định kỳ về các hoạt động của mình, cũng như các nghĩa vụ khác để ngăn chặn các giao dịch vi phạm pháp luật Việt Nam.

### **1.3.2 Cơ sở hạ tầng kỹ thuật**

Hệ thống kết cấu hạ tầng cho phát triển kinh tế số ở Việt Nam trong thời gian qua được tập trung đầu tư xây dựng, ngày càng đồng bộ, hiện đại, cụ thể là hạ tầng công nghệ thông tin, viễn thông quốc gia được phát triển mạnh, phủ sóng rộng khắp. Mạng di động ở Việt Nam phát triển mạnh, tỷ lệ phủ sóng đạt 99,7%. Hệ thống mạng di động 3G, 4G đã được phủ sóng rộng khắp. Tỷ lệ người sử dụng Internet ở Việt Nam đã vượt mức trung bình của khu vực và thế giới (Năm 2016 có 46,55% dân số Việt Nam truy cập Internet; năm 2018 có 64 triệu người ở Việt Nam sử dụng Internet). Việt Nam cũng từng bước phát triển công nghệ vệ tinh. Đến đầu năm 2019, Việt Nam đã phóng một số vệ tinh, trong đó có vệ tinh nhằm mục đích cung cấp Internet cho vùng sâu, vùng xa.

### **1.3.3 Cơ sở hạ tầng thanh toán**

Hình thức thanh toán điện tử phổ biến tại Việt Nam: thanh toán điện tử bằng thẻ ngân hàng, thanh toán qua cổng thanh toán, thanh toán bằng ví điện tử, thanh toán bằng di động.

### **1.3.4 Bảo mật trong thương mại điện tử**

Sử dụng https, bảo mật server và admin panel, bảo mật hệ thống thanh toán, sử dụng phần mềm diệt virus và malware (end-point security), chứng chỉ ssl nâng cao (ev ssl), plugin bảo mật dành cho tmĐT

## **1.4 Ứng dụng mô hình học sâu vào bài toán nhận dạng hình ảnh trong thương mại điện tử**

Với sự phát triển mạnh mẽ của TMĐT đồng nghĩa với việc các mặt hàng sản phẩm được bày bán trên các trang thương mại điện tử ngày càng nhiều. Số lượng và chủng loại nhiều như vậy thì việc nhận dạng các sản phẩm là rất cần thiết. Điều này phục vụ cho việc tìm kiếm, gợi ý sản phẩm cho khách hàng. Hiện nay, các mô hình học sâu hay còn gọi là Deep Learning có khả năng nhận dạng, phân loại hình ảnh cực kì tốt. Với sự hiệu quả của việc phân loại hình ảnh từ mô hình học sâu, thì hình ảnh các sản phẩm thương mại điện tử cũng tương tự được phân loại tốt. Việc tìm kiếm hay gợi ý quảng cáo sản phẩm sẽ trở lên chính xác hơn. Khách hàng sẽ ưng ý hơn với các kết quả hiển thị sản phẩm để mua hàng một cách tiện lợi.

## **1.5 Kết luận chương I**

Trong chương I này, luận văn đã giới thiệu tổng quan về thương mại điện tử trên thế giới và ở Việt Nam. Làm rõ các thực trạng của thương mại điện tử ở nước ta và trên thế giới. Trình bày về hạ tầng thương mại điện tử ở nước ta bao gồm về: pháp lý, kỹ thuật, thanh toán và bảo mật. Đồng thời cũng nêu khát quát ứng dụng mô hình học sâu vào bài toán nhận dạng hình ảnh trong thương mại điện tử.

## CHƯƠNG II: GIỚI THIỆU VÀ NGHIÊN CỨU CNN CHO BÀI TOÁN NHẬN DIỆN HÌNH ẢNH

### 2.1 Khát quát bài toán nhận diện hình ảnh và CNN

#### 2.1.1 Bài toán nhận diện hình ảnh

Phân loại hình ảnh (Image classification) hay Nhận dạng hình ảnh (Image recognition) là một trong những tác vụ của thị giác máy tính, ở đó thuật toán xem xét và dán nhãn cho hình ảnh từ một tập danh mục được xác định và đào tạo trước.

#### 2.1.2 Học máy, học sâu và CNN

##### 2.1.2.1 Học máy

Học máy hay còn gọi với cái tên Tiếng Anh là Machine Learning. Tom Mitchell đưa ra một định nghĩa hiện đại và rõ ràng:

“Một chương trình máy tính được cho là học hỏi từ kinh nghiệm E đối với một số loại nhiệm vụ T và thước đo hiệu suất P, nếu hiệu suất của nó ở các nhiệm vụ trong T, được đo bằng P, cải thiện theo kinh nghiệm E.”

##### 2.1.2.2 Học sâu

- **Mạng nơron nhân tạo (Neural Network)**

Mạng nơron nhân tạo là một hệ thống học tập tính toán sử dụng một mạng lưới các chức năng để hiểu và dịch đầu vào của dữ liệu để thành đầu ra mong muốn, thường là ở dạng khác.

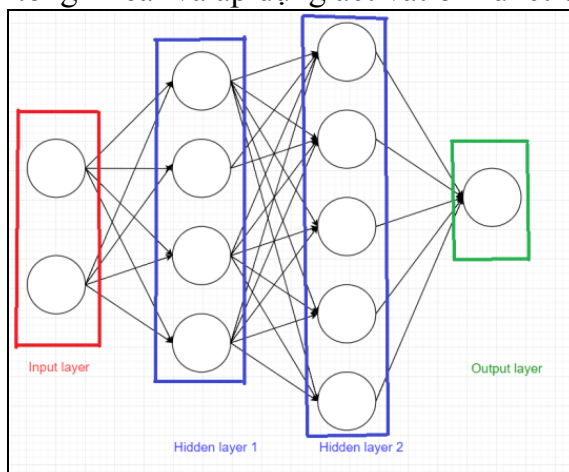
- **Mô hình tổng quát**

Kiến trúc mạng Neural Network thường bao gồm 3 kiểu thành phần chính:

- -Input layer (lớp đầu vào): Lớp này nằm bên trái cùng của mạng, thể hiện cho các đầu vào của mạng.
- -Output layer (lớp đầu ra): Là lớp bên phải cùng và nó thể hiện cho những đầu ra của mạng.
- Hidden layer (lớp ẩn): Lớp này nằm giữa lớp đầu vào vào và lớp đầu ra nó thể hiện cho quá trình suy luận logic của mạng.

Các hình tròn được gọi là node. Mỗi node trong hidden layer và output layer :

- Liên kết với tất cả các node ở layer trước đó với các hệ số  $w$  riêng.
- Mỗi node có 1 hệ số bias  $b$  riêng.
- Diễn ra 2 bước: tính tổng linear và áp dụng activation function.

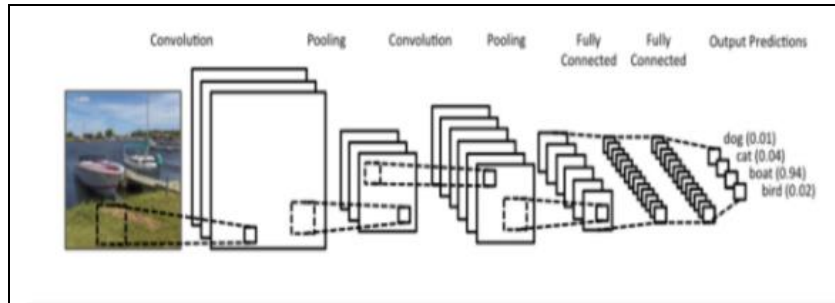


Hình 1. Mô hình mạng neural network



### 2.1.2.3 Mạng nơ ron tích chập (CNN)

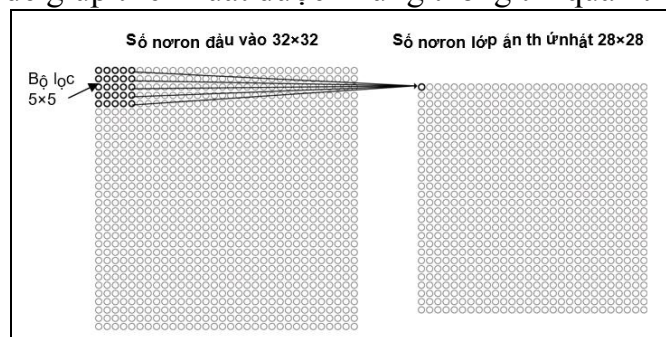
Convolutional Neural Network (CNN – Mạng nơron tích chập) là một kiểu mạng NN truyền thẳng, trong đó kiến trúc chính gồm nhiều thành phần được ghép nối với nhau theo cấu trúc nhiều tầng bao gồm: lấy chập (Convolution), lấy gộp (Pooling), kích hoạt phi tuyến (Non-Linear Activation) và kết nối đầy đủ (Fully-Connected).



Hình 2. Mô hình mạng neural tích chập

- **Lấy chập (Convolution)**

Là thao tác đầu tiên quan trọng nhất trong cấu trúc của mạng học sâu CNN. Phép lấy chập dựa trên lý thuyết xử lý tín hiệu số, thực hiện các xử lý về mặt toán học tính lấy chập để giúp trích xuất được những thông tin quan trọng từ dữ liệu.



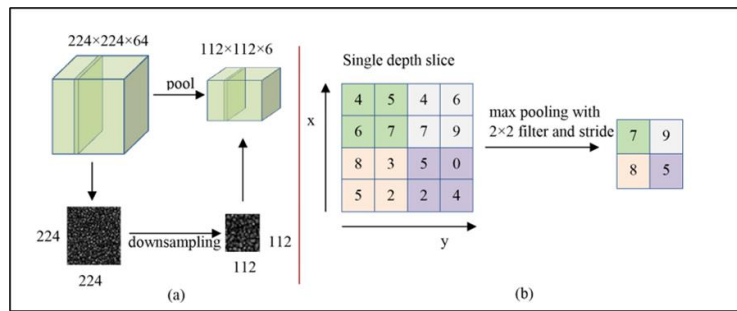
Hình 3. Mô tả lấy chập dùng bộ lọc kích thước 5x5

- **Kích hoạt phi tuyến (Non-Linear Activation)**

Về cơ bản, lấy chập là một phép biến đổi tuyến tính. Nếu tất cả các nơron được tổng hợp bởi các phép biến đổi tuyến tính thì một mạng nơron đều có thể đưa về dưới dạng một hàm tuyến tính. Khi đó mạng NN sẽ đưa các bài toán về hồi qui logistic (logistic regression). Do đó, sau mỗi lớp lấy chập, đầu ra của ảnh xạ lấy chập được cho qua hàm kích hoạt phi tuyến.

- **Lấy gộp (Pooling)**

Lấy gộp (pooling) hay còn gọi subsampling hoặc downsampling là một trong những thành phần tính toán chính trong cấu trúc CNN. Xét về mặt toán học, lấy gộp thực chất là quá trình tính toán trên ma trận đầu vào trong đó mục tiêu đạt được sau khi tính toán là giảm kích thước ma trận nhưng vẫn làm nổi bật lên được đặc trưng có trong ma trận đầu vào. Trong CNN, toán tử pooling được thực hiện độc lập trên mỗi kênh màu của ma trận ảnh đầu vào. Có nhiều toán tử pooling như sum-pooling, max-pooling, L2-pooling song max-pooling thường được sử dụng. Về mặt ý nghĩa, max-pooling xác định vị trí cho tín hiệu mạnh nhất khi áp dụng một loại bộ lọc.



Hình8. Ví dụ sử dụng max-pooling

- **Kết nối đầy đủ (Fully-Connected)**

Là cách kết nối các neuron ở hai tầng với nhau trong đó tầng sau kết nối đầy đủ với các neuron ở tầng trước nó. Đây cũng là dạng kết nối thường thấy ở NN. Trong CNN, tầng này thường được sử dụng ở các tầng phía cuối của kiến trúc mạng kết nối với đầu ra của mạng. Lớp này cơ bản là lấy thông tin đầu vào (có thể là đầu ra của lớp lấy chập hoặc kích hoạt phi tuyến hoặc lấy gộp) còn đầu ra là vectơ  $N$  chiều với  $N$  là số lớp cần phân lớp.

## 2.2 Một số nghiên cứu có liên quan

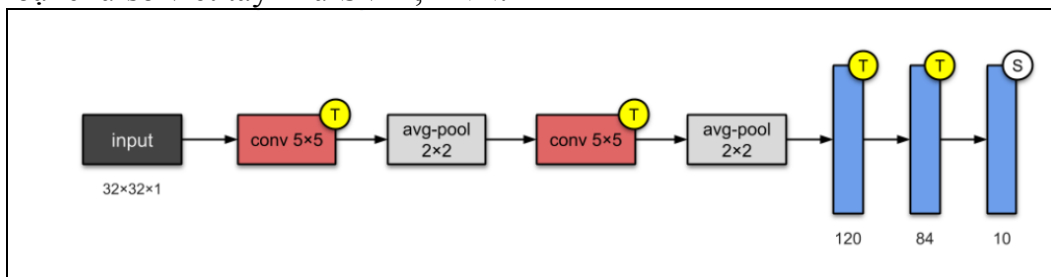
### 2.2.1 Quá trình phát triển của các kiến trúc CNN

- 1998: Yan Lecun lần đầu tiên sử dụng mạng tích chập trong tác vụ phân loại chữ số viết tay và đạt hiệu quả cao.
- 2009: Bộ dữ liệu ImageNet được giới thiệu vào năm 2009 là một trong những bộ dữ liệu tạo ra sự thay đổi trong cộng đồng computer vision. Kể từ thời điểm 2010, ImageNet trở thành tiêu chuẩn đo đặc sự phát triển của các thuật toán học có giám sát trong thị giác máy tính.
- 2012: Mạng AlexNet sử dụng tích chập CNN lần đầu tiên vượt qua các phương pháp tạo đặc trưng thủ công truyền thống như HOG, SHIFT và đạt độ chính xác cách biệt trong cuộc thi ImageNet.

### 2.2.2. Các mạng CNN tiêu biểu

- **LeNet-5 (1998)**

LeNet-5 là kiến trúc đầu tiên áp dụng mạng tích chập 2 chiều của giáo sư Yan Lecun, cha đẻ của kiến trúc CNN. Model ban đầu khá đơn giản và chỉ bao gồm 2 convolutional layers + 3 fully-connected layers. Mặc dù đơn giản nhưng nó có kết quả tốt hơn so với các thuật toán machine learning truyền thống khác trong phân loại chữ số viết tay như SVM, kNN.



Hình 4. Kiến trúc LeNet

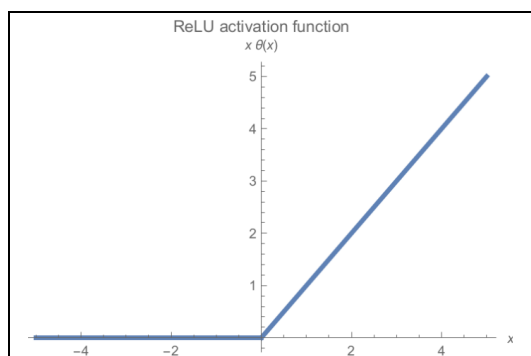
- **AlexNet (2012)**

Ý tưởng của AlexNet dựa trên LeNet của Yan Lecun và cải tiến ở các điểm:

- Tăng kích thước đầu vào và độ sâu của mạng.
- Sử dụng các bộ lọc (kernel hoặc filter) với kích thước giảm dần qua các layers để phù hợp với kích thước của đặc trưng chung và đặc trưng riêng.
- Sử dụng local normalization để chuẩn hóa các layer giúp cho quá trình hội tụ nhanh hơn.

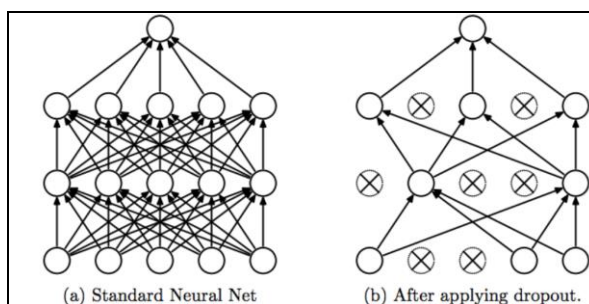
Ngoài ra mạng còn cải tiến trong quá trình optimizer như:

- Lần đầu tiên sử dụng activation là ReLU (Rectified Linear Unit) thay cho Sigmoid.



Hình 5. Hàm ReLu

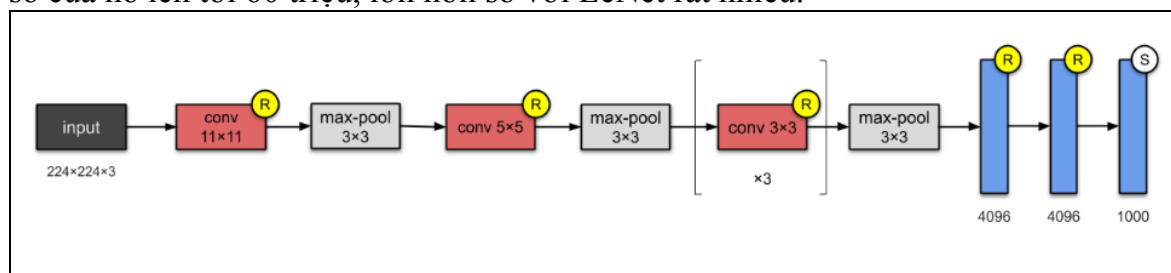
- Sử dụng dropout layer giúp giảm số lượng liên kết neural và kiểm soát overfitting.



Hình 6. Minh họa phương pháp dropout

- Qua các layers, kích thước output giảm dần nhưng độ sâu tăng dần qua từng kernel.

Mạng AlexNet có resolution của input và số lượng layer lớn hơn nên số lượng tham số của nó lên tới 60 triệu, lớn hơn so với LeNet rất nhiều.

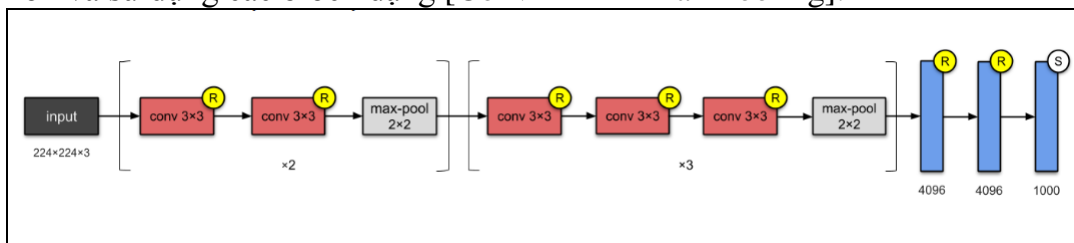


Hình 7.. Kiến trúc mạng AlexNet

- **VGG-16 (2014)**

Với VGG-16, quan điểm về một mạng nơ ron sâu hơn sẽ giúp ích cho cải thiện độ chính xác của mô hình tốt hơn. Về kiến trúc thì VGG-16 vẫn giữ các đặc điểm của AlexNet nhưng có những cải tiến.

Bắt đầu từ VGG-16, một hình mẫu chung cho các mạng CNN trong các tác vụ học có giám sát trong xử lý ảnh đã bắt đầu hình thành đó là các mạng trở nên sâu hơn và sử dụng các block dạng [Conv2D\*n + Max Pooling].



Hình 8. Kiến trúc VGG-16

## 2.3 Học chuyển giao

Các mô hình nổi tiếng, được train trên các dataset lớn (MNIST, CIFAR-100, ImageNet, ...) và source code cũng như các trọng số(weights) của model được công khai cho cộng đồng. Chúng ta gọi những Model đi kèm Weights như vậy là một Pretrained Model. Model mới sử dụng một phần hay toàn bộ pretrained model như một phần của nó để học một nhiệm vụ mới được gọi là Transferred Model.

Có 2 loại Transfer Learning:

- Feature extractor: Sử dụng transfer learning chỉ để trích xuất đặc trưng rồi sau đó sử dụng những vector đặc trưng ấy vào mục đích khác.
- Fine tuning: Thường bỏ đi lớp đầu ra cuối cùng (output layer) hoặc một phần các lớp trước đó của của ConvNet. Sau đó thêm các lớp tích chập, lớp gộp, lớp kết nối đầy đủ, lớp đầu ra vào sau, tạo thành một mạng CNN mới.

## 2.4 Mô hình kiến trúc ResNet và áp dụng vào bài toán phân loại ảnh

### 2.4.1. Giới thiệu về ResNet

ResNet là mạng CNN nổi bật có cả kích thước mô hình và độ chính xác đều lớn hơn MobileNet. Nó được ra mắt lần đầu tiên vào năm 2015 trong một bài báo Nghiên cứu thực tế sâu về nhận dạng hình ảnh và rất nhanh chóng đạt được thứ hạng đầu tiên trên ILSVLC 2015. Nó cho phép bạn điều chỉnh độ sâu của mô hình theo yêu cầu của mình một cách linh hoạt nhất có thể. Vì vậy nhiều loại phiên bản sâu ResNet như ResNet18, ResNet34, ResNet50, ResNet101, ResNet152

Đặc điểm cụ thể nhất trong ResNet là kết nối tắt được áp dụng bên trong mỗi khối. Như vậy để giúp mô hình giữ được đặc trưng từ trước và sau. Do đó, ResNet là viết tắt của Residual Learning Network.

### 2.4.2. Batch Normalization

ResNet là kiến trúc đầu tiên được áp dụng Batch Normalization bên trong mỗi khối Residual trên cơ sở rằng mô hình có thể hay bị hiện tượng vanishing gradient descent khi nó sâu hơn. Batch Normalization giúp giữ ổn định trên dọc xuống và hỗ trợ quá trình đào tạo hội tụ nhanh chóng đến điểm tối ưu.

Batch Normalization được áp dụng trên mỗi mini-batch bởi standard normalization  $N(0,1)$ . Ví dụ, chúng ta có  $\mathbf{B} = \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m$  với  $m$  là kích thước của mini-batch size.

Tất cả các mẫu dữ liệu đầu vào sẽ được chia tỷ lệ lại như dưới đây:

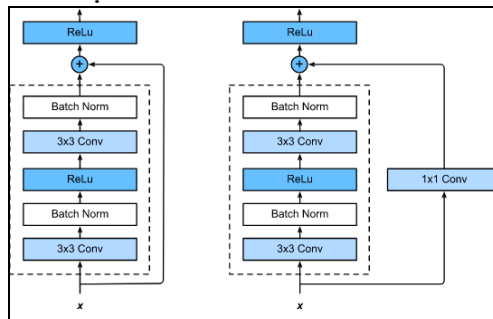
$$\hat{x}_i = \frac{x_i - \mu}{\sigma}$$

Với các giá trị  $\mu$  và  $\sigma$  được tính như sau:

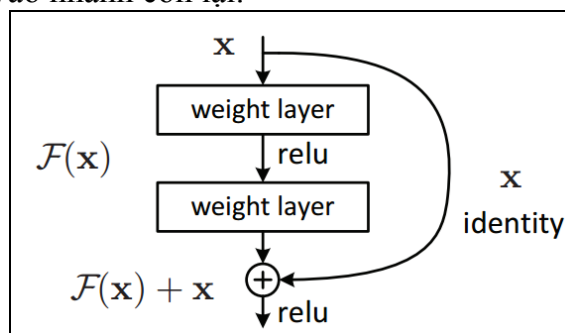
$$\mu = \frac{1}{m} \sum_{i=1}^m x_i \quad \sigma^2 = \sum_{i=1}^m (x_i - \mu)^2$$

### 2.4.3 Kết nối tắt (Skip Connection)

ResNet có khối tích chập (Convolutional Block, chính là Conv block trong hình) sử dụng bộ lọc kích thước  $3 \times 3$  giống với của InceptionNet. Khối tích chập bao gồm 2 nhánh tích chập trong đó một nhánh áp dụng tích chập  $1 \times 1$  trước khi cộng trực tiếp vào nhánh còn lại.



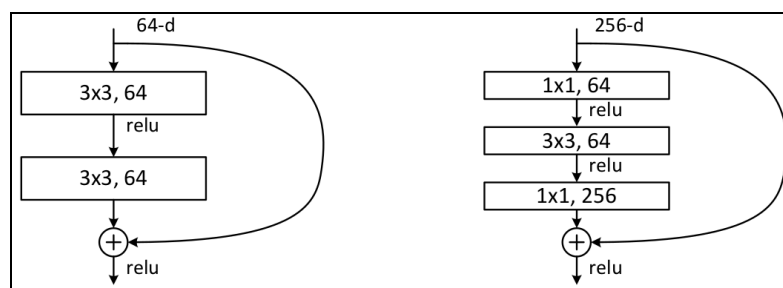
Hình 9. Khối ResNet thông thường và khối ResNet với tầng tích chập  $1 \times 1$   
Khối xác định (Identity block) thì không áp dụng tích chập  $1 \times 1$  mà cộng trực tiếp giá trị của nhánh đó vào nhánh còn lại.



Hình 10. Khối xác định (Identity block)

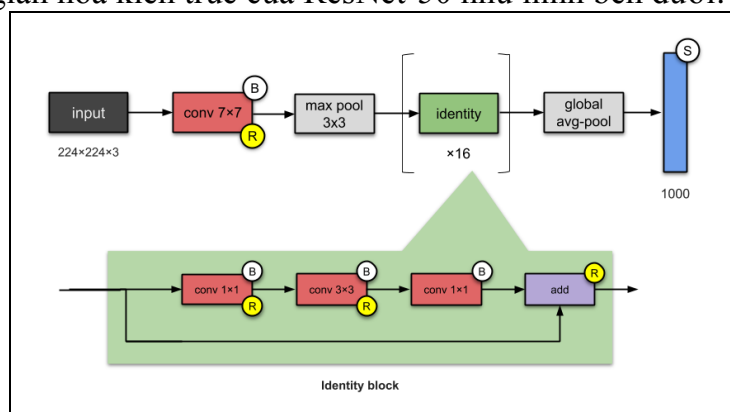
### 2.4.4. Mô hình ResNet50 áp dụng vào bài toán

Kiến trúc Resnet50 cũng giống như kiến trúc ResNet ở trên, có một điểm khác biệt chính. So với Resnet34, các kết nối tắt đã qua hai lớp nhưng bây giờ chúng qua ba lớp và cũng có các lớp tích chập  $1 \times 1$  được thêm vào. Điều này có độ chính xác cao hơn nhiều so với mô hình ResNet 34 lớp. ResNet 50 lớp đạt được hiệu suất 3,8 tỷ FLOPS.



Hình 11. Kết nối tắt qua 3 lớp của ResNet50 so với 2 lớp của ResNet34

Kiến trúc cơ sở của ResNet50 chỉ gồm các khối tích chập và khối xác định. Ta có thể đơn giản hóa kiến trúc của ResNet-50 như hình bên dưới:



Hình 12. Kiến trúc tóm tắt của mạng ResNet50.

Lý do chọn lựa mô hình ResNet50 để sử dụng vào bài toán nhận dạng hình ảnh sản phẩm thương mại điện tử là kiến trúc được sử dụng phổ biến nhất ở thời điểm hiện tại. Mặc dù là một mạng rất sâu khi có số lượng layer lên tới 152 nhưng nhờ áp dụng những kỹ thuật kết nối tắt đặc biệt mà đã trình bày ở trên nên kích thước của ResNet50 chỉ khoảng 26 triệu tham số. Điều đó khiến cho mô hình ResNet50 trở lên hiệu quả, rất phù hợp để nhận dạng hình ảnh sản phẩm thương mại điện tử với số loại sản phẩm cần nhận dạng là 10 loại.

Áp dụng mô hình ResNet50 sẽ là phương pháp giải quyết bài toán nhận dạng hình ảnh. Kết hợp cùng với học chuyển giao (transfer learning) đã đề cập ở trên, ta chỉ việc thay lớp đầu ra phân loại mặc định của mô hình là 1000 bằng số lớp cần phân loại trong tập ảnh dữ liệu. Sử dụng pretrained model ResNet50 của ImageNet dataset. Bộ dữ liệu ImageNet được giới thiệu vào năm 2009 là một trong những bộ dữ liệu tạo ra sự thay đổi trong cộng đồng computer vision. Đây là bộ dữ liệu lớn nhất so với các bộ dữ liệu từng có từ trước đến thời điểm đó. Với kích thước lên tới 1 triệu ảnh và phân bố đều trên 1000 nhãn.

## 2.5 Kết luận chương II

Chương II của luận văn đã nêu khái quát về bài toán nhận dạng hình ảnh. Làm rõ các khái niệm học máy và các loại thuật toán của học máy và có thể thấy rằng bài toán nhận dạng hình ảnh thuộc nhóm thuật toán phân loại của học máy. Luận văn cũng đã viết chi tiết về học sâu. Trình bày cụ thể về mạng nơ ron tích chập CNN và thành phần cụ thể của mạng này. Nêu ra quá trình phát triển của mạng

CNN và trình bày kiến trúc một số mạng tiêu biểu như LeNet-5, AlexNet, VGG16. Trình bày về học chuyển giao giúp tăng hiệu quả cho các mô hình học sâu. Luận văn trình bày cụ thể về kiến trúc mạng được sử dụng phổ biến và rộng rãi hiện nay là ResNet, viết chi tiết về các thành phần của mạng này: Batch Normalization, kết nối tắt. Đồng thời trình bày kiến trúc mạng ResNet50 kết hợp học chuyển giao sử dụng pretrained model ResNet50 của ImageNet dataset để phân loại hình ảnh.

Tại chương III dưới đây sẽ trình bày rõ hơn về thực nghiệm sử dụng model ResNet50 trong nhận dạng hình ảnh.



## CHƯƠNG III: THỰC NGHIỆM, ĐÁNH GIÁ VÀ XÂY DỰNG HỆ THỐNG NHẬN DẠNG HÌNH ẢNH TRONG THƯƠNG MẠI ĐIỆN TỬ

### 3.1 Dữ liệu thử nghiệm

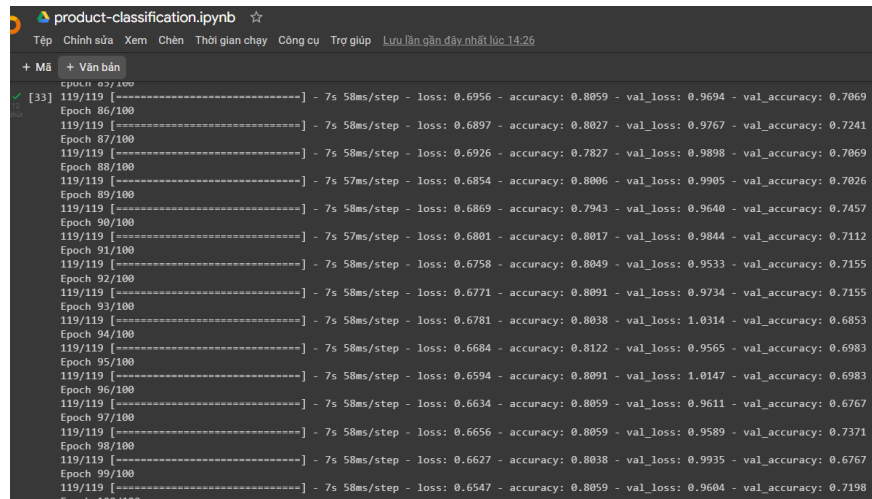
Tập ảnh thử nghiệm là tập ảnh các sản phẩm thương mại điện tử phổ biến gồm các loại mặt hàng khác nhau. Tập dữ liệu gồm 10 sản phẩm, thuộc 3 loại sản phẩm: đồ điện tử, đồ gia dụng, đồ thời trang.

Tập ảnh có kích thước dung lượng là 24.8 MB. Trong đó gồm ảnh các sản phẩm thương mại điện tử với 1182 ảnh chia vào 2 thư mục train và test. Trong đó tập test chiếm 20% tổng số ảnh.

### 3.2 Tiến hành thực nghiệm và kết quả

#### 3.2.1 Tiến hành thực nghiệm

Tiến hành đào tạo mô hình ResNet50 cho tập dữ liệu hình ảnh các sản phẩm thương mại điện tử. Đào tạo mô hình qua 100 epochs, với mỗi một epoch sau khi training xong thì sẽ hiện ra kết quả hàm mất mát và độ chính xác của mô hình trên cả tập đào tạo (train) và tập kiểm tra (val). Điều này giúp ta trong quá trình đào tạo mô hình có thể thấy được độ quá trình học của mô hình và hiệu quả của mô hình sau các epoch.



Hình 13. Các epochs trong quá trình đào tạo mô hình

#### 3.2.2. Kết quả thực nghiệm

- Kết quả dự đoán của một số ảnh mặt hàng thuộc loại sản phẩm đồ điện tử:



Hình 14. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 1

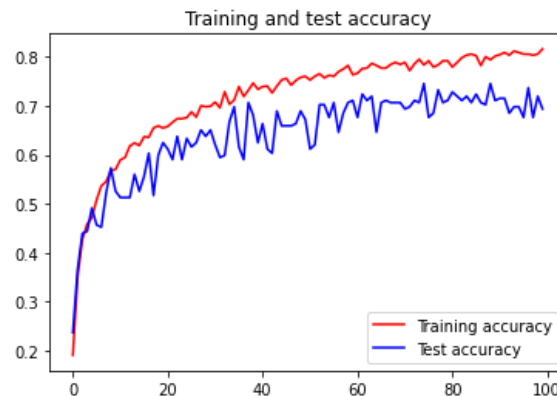


### 3.3 Đánh giá kết quả

#### 3.3.1. Đánh giá bằng độ Accuracy

Chúng ta sẽ đánh giá kết quả thực nghiệm nhận dạng các sản phẩm thương mại điện tử bằng mô hình ResNet50 với pretrained imagenet bằng độ đánh giá Accuracy. Độ chính xác Accuracy của phân loại là số liệu tóm tắt hiệu suất của mô hình phân loại dưới dạng số dự đoán đúng chia cho tổng số dự đoán. Mô hình cho độ chính xác khá tốt: 74.36%

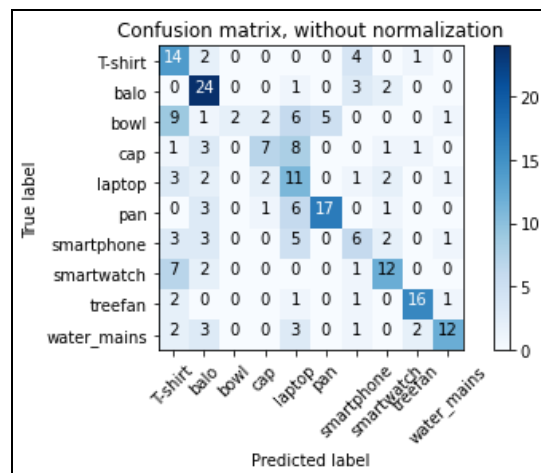
Nhìn vào hình 17 vẽ biểu đồ đường của training accuracy và test accuracy, lúc đầu độ chính xác tăng lên rất nhanh nhưng càng về sau càng chậm. Và qua 100 epochs chúng tạo thành một đường cong.



Hình 17. Biểu đồ đường độ chính xác accrancy qua các epochs

#### 3.3.2. Đánh giá bằng Confusion matrix

Dưới đây là confusion matrix của mô hình nhận dạng sản phẩm thương mại điện tử trên tập test:



Hình 18. Confusion matrix của mô hình trên tập test

Nhìn vào confusion matrix ở trên của có thể thấy được các sản phẩm được nhận dạng chính xác nhất, nhận dạng kém chính xác nhất, sản phẩm nào hay bị nhận dạng nhầm thành sản phẩm nào.

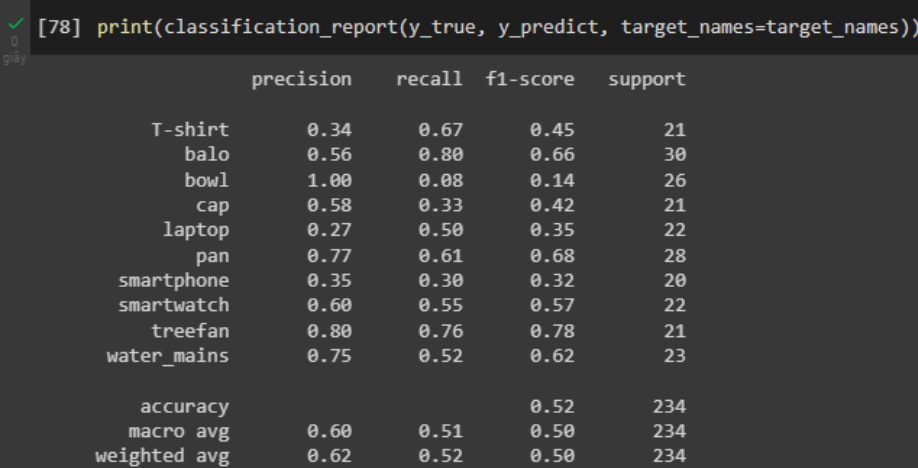
Nhận xét:

- Sản phẩm được nhận dạng tốt nhất là Balo với 24/30 hình ảnh Balo được nhận dạng chính xác.

- Sản phẩm bị nhận dạng kém nhất là Bowl với 2/26 hình ảnh Bowl được nhận dạng chính xác. Lốp Bowl hay bị nhận nhầm thành T-shirt (9/26).
- Một số sản phẩm khác cũng cho kết quả nhận dạng tốt, ít bị nhầm thành sản phẩm khác như T-shirt, Pan, Smartphone, Smartwatch, Water\_mains.
- Các hình ảnh sản phẩm nhận dạng sai hay bị nhận nhầm thành hai mặt hàng T-shirt và Laptop.

### 3.2.2. Đánh giá bằng Precision, Recall, F1-Score

Sử dụng hàm `classification_report()` đánh giá mô hình trên tập test ta sẽ thấy được các thông số phân loại Precision, Recall, F1-Score của từng sản phẩm và Macro-F1 Score, Micro-F1 Score của chung 10 sản phẩm. Trong `classification_report()` thì Micro-F1 Score sẽ được ghi là accuracy.



```
[78] print(classification_report(y_true, y_predict, target_names=target_names))
```

	precision	recall	f1-score	support
T-shirt	0.34	0.67	0.45	21
balo	0.56	0.80	0.66	30
bowl	1.00	0.08	0.14	26
cap	0.58	0.33	0.42	21
laptop	0.27	0.50	0.35	22
pan	0.77	0.61	0.68	28
smartphone	0.35	0.30	0.32	20
smartwatch	0.60	0.55	0.57	22
treefan	0.80	0.76	0.78	21
water_mains	0.75	0.52	0.62	23
accuracy			0.52	234
macro avg	0.60	0.51	0.50	234
weighted avg	0.62	0.52	0.50	234

Hình 19. Precision, Recall, F1-Score của mô hình trên tập test  
Nhận xét:

- Sản phẩm có F1-Score cao nhất là Treefan bằng 0.78 với Precision là 0.8 và Recall là 0.76.
- Sản phẩm có F1-Score thấp nhất là Bowl bằng 0.14 với Precision là 1 và Recall là 0.08, chênh lệch nhau quá lớn. Nhìn vào chỉ số như vậy có thể thấy rằng các ảnh Bowl thật sự bị nhận nhầm thành ảnh khác khá nhiều. Tuy nhiên Precision = 1 cho thấy ảnh nào được nhận dạng là Bowl thì chắc chắn đúng.
- Sản phẩm có Recall cao nhất là Balo với 0.8, khá cao. Có rất ít ảnh Balo bị bỏ sót nhận nhầm thành ảnh nhãn khác.
- Có 4/10 sản phẩm có F1 Score > 0.6 và 5/10 sản phẩm F1-Score > 0.5.
- Macro Precision = 0.6 cho toàn bộ 10 sản phẩm và Macro Precision = 0.51, ta được Macro-F1 Score bằng 0.5.
- Macro-F1 Score của toàn bộ 10 sản phẩm là 0.52.
- Nhìn chung, mô hình cho các chỉ số khá ổn.

## 3.4 Áp dụng xây dựng hệ thống

### 3.4.1 Xây dựng web nhận dạng sản phẩm

Từ mô hình RestNet50 đã được đào tạo ở trên, ta xây dựng một trang web nhận dạng các hình ảnh sản phẩm của thương mại điện tử. Công nghệ để xây dựng ứng dụng website sử dụng framework Flask. Ngôn ngữ lập trình là Python và dùng IDE Visual Studio Code để chạy ứng dụng.

### 3.4.2 Tạo web với Plask

Flask là một Web Framework rất nhẹ của Python, dễ dàng giúp người mới bắt đầu học Python có thể tạo ra website nhỏ. Flask cũng dễ mở rộng để xây dựng các ứng dụng web phức tạp. Flask có nền tảng là Werkzeug WSGI và Jinja2 và nó đã trở thành một trong những Web Framework phổ biến nhất của Python.

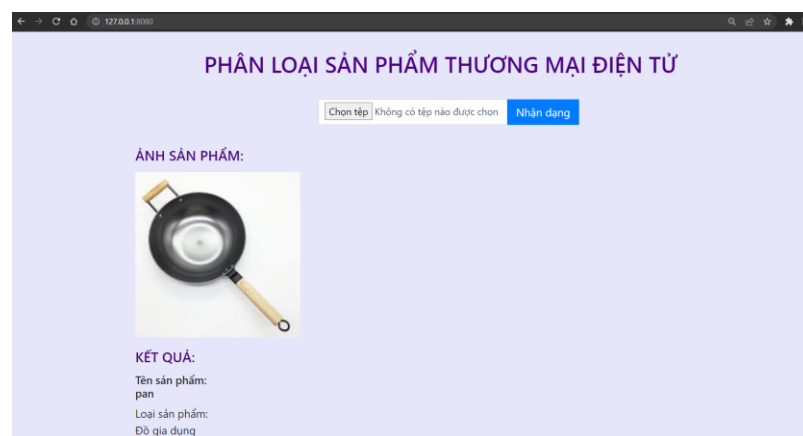
### 3.4.3 Trang web phân loại sản phẩm thương mại điện tử

Trong trang web phân loại sản phẩm thương mại điện tử thì sử dụng chủ yếu 2 API RESTful là GET và POST của framework Flask. Mục đích là để POST ảnh truy vấn và GET ảnh kết quả tìm kiếm và các thông số liên quan.

### 3.4.4 Trang web nhận dạng sản phẩm thương mại điện tử

Sau khi có giao diện hoàn chỉnh, kết hợp cùng với mô hình RestNet50 đã được đào tạo ở trên. Ta được website nhận dạng sản phẩm thương mại điện tử có chức năng nhận diện phân loại được ảnh sản phẩm thương mại điện tử mà người dùng tải ảnh lên. Giao diện trang web có một nút “Chọn tệp” để upload file ảnh và một nút “Nhận dạng” để nhận dạng sản phẩm.

Dưới đây là giao diện của trang web nhận dạng sản phẩm thương mại điện tử. Trang web có giao diện đơn giản về bố cục, dễ sử dụng cho tất cả mọi người.



Hình 1. Web hiển thị ảnh sản phẩm và kết quả nhận dạng

## 3.5 Kết luận chương III

Trong chương này, luận văn đã tiến hành thực nghiệm mô hình RestNet50 với tập dữ liệu hình ảnh các sản phẩm thương mại điện tử. Mô hình RestNet50 sau khi được đào tạo cho độ chính xác 74.36% trên tập kiểm thử. Xây dựng trang web nhận dạng hình ảnh sản phẩm thương mại điện tử sử dụng mô hình học sâu ResNet50 đã được huấn luyện ở trên. Khi ảnh sản phẩm được người dùng tải lên trang web thì trang web sẽ hiển thị kết quả nhận dạng sản phẩm và loại sản phẩm thương mại điện tử.

## KẾT LUẬN

### 1. Kết quả đạt được:

- Luận đã trình bày chi tiết về mặt khái niệm, hình thức, đặc trưng của thương mại điện tử và ưu nhược điểm và tác động của nó đến doanh nghiệp.
- Luận văn nêu ra thực trạng của thương mại điện tử ở nước ta và trên thế giới.
- Trình bày về hạ tầng thương mại điện tử ở nước ta bao gồm về: pháp lý, kỹ thuật, thanh toán và bảo mật.
- Luận văn nêu khái quát về bài toán nhận dạng hình ảnh.
- Nêu khái niệm học máy và các loại thuật toán của học máy và đã viết chi tiết về học sâu.
- Trình cụ thể về mạng nơ ron tích chập CNN và thành phần cụ thể của mạng này. Nêu ra quá trình phát triển của mạng CNN và trình bày kiến trúc một số mạng tiêu biểu. Trình bày về học chuyển giao giúp tăng hiệu quả cho các mô hình học sâu.
- Luận văn viết cụ thể về kiến trúc mạng được sử dụng phổ biến và rộng rãi hiện nay là ResNet
- Trình bày kiến trúc mạng ResNet50 kết hợp học chuyển giao sử dụng pretrained model ResNet50 của ImageNet dataset để phân loại hình ảnh.
- Tiến hành đào tạo cho mô hình ResNet50 sử dụng framework Tensorflow. Mô hình ResNet50 sau khi được đào tạo cho độ chính xác 74.36% trên tập kiểm thử.
- Xây dựng trang web nhận diện sản phẩm thương mại điện tử sử dụng mô hình ResNet50 đã được huấn luyện với tập thử nghiệm.

### 2. Hướng nghiên cứu tiếp theo:

- Với những kết quả trên, hướng nghiên cứu tiếp theo sẽ nghiên cứu sâu hơn về các mô hình học sâu nhận dạng ảnh kiểu mới, cho kết quả chính xác hơn. Ứng dụng nhận dạng hình ảnh nhiều hơn để thêm tính năng trong các hệ thống thương mại điện tử.