

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**

LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

HÀ NỘI - 2022

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

LUẬN VĂN THẠC SĨ KỸ THUẬT

(Theo định hướng ứng dụng)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. NGUYỄN VĂN THỦY

HÀ NỘI 2022

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn “Mô hình học sâu và ứng dụng cho bài toán nhận dạng hình ảnh trong thương mại điện tử” là một công trình nghiên cứu của tôi dưới sự hướng dẫn của giảng viên hướng dẫn, không sao chép lại của người khác. Các tài liệu được luận văn tham khảo, kế thừa và trích dẫn đều được liệt kê trong danh mục các tài liệu tham khảo.

Tôi xin chịu hoàn toàn trách nhiệm về lời cam đoan nêu trên.

Hà Nội, ngày tháng năm 2022

Học viên

Bùi Việt Anh

LỜI CẢM ƠN

Đầu tiên em xin cảm ơn đến các thầy cô giảng viên của Học viện Công nghệ Bưu chính Viễn thông đã truyền đạt cho em bao kiến thức vô cùng quý báu và cần thiết trong thời gian học tập ở trường. Những tri thức ấy chính là nền tảng vững chắc cho sự phát triển của em sau này. Xin kính chúc thầy cô có nhiều sức khỏe và thành công hơn nữa trong sự nghiệp trồng người.

Em cũng xin gửi lời cảm ơn Tiến sĩ Nguyễn Văn Thủy, giảng viên đã tận tình hướng dẫn em trong suốt quá trình nghiên cứu để hoàn thành đồ án này. Nhờ sự hướng dẫn chỉ bảo tận tình của thầy, em đã có thêm nhiều kiến thức về trí tuệ nhân tạo, học máy và xử lý ảnh. Vốn kiến thức quý giá vô cùng quan trọng cho định hướng của em ở tương lai.

Qua những năm tháng học tập tại trường, em đã gặp vô vàn khó khăn nhưng thật may mắn khi gia đình và thầy cô, bạn bè luôn ở bên và động viên giúp đỡ. Em xin gửi lời cảm ơn tới tất cả mọi người.

Em xin chân thành cảm ơn !

MỤC LỤC

LỜI CAM ĐOAN	i
LỜI CẢM ƠN	ii
DANH MỤC TỪ VIẾT TẮT.....	vi
DANH MỤC BẢNG BIỂU/HÌNH VẼ.....	viii
MỞ ĐẦU	1
1. Lý do chọn đề tài	1
2. Tổng quan về vấn đề nghiên cứu.....	1
3. Mục đích nghiên cứu	2
4. Đối tượng và phạm vi nghiên cứu	2
5. Phương pháp nghiên cứu	3
CHƯƠNG I: TỔNG QUAN VỀ THƯƠNG MẠI ĐIỆN TỬ	4
1.1 Giới thiệu về thương mại điện tử.....	4
1.1.1 Lịch sử phát triển	4
1.1.2 Khái niệm.....	4
1.1.3 Cách thức hoạt động	5
1.1.4 Các hình thức.....	6
1.1.5 Đặc trưng	8
1.1.6 Ưu và nhược điểm của thương mại điện tử	10
1.1.7 Lợi ích thương mại điện tử đến doanh nghiệp	12
1.2 Thực trạng thương mại điện tử trên thế giới và Việt Nam	13
1.2.1 Thương mại điện tử toàn cầu.....	13
1.2.2 Thực trạng phát triển thương mại điện tử tại Việt Nam	15
1.3 Hạ tầng thương mại điện tử ở Việt Nam hiện nay.....	17
1.3.1 Cơ sở hạ tầng pháp lý	17
1.3.2 Cơ sở hạ tầng kỹ thuật	18
1.3.3 Cơ sở hạ tầng thanh toán.....	19
1.3.4 Bảo mật trong thương mại điện tử.....	21

1.4 Ứng dụng mô hình học sâu vào bài toán nhận dạng hình ảnh trong thương mại điện tử	22
1.5 Kết luận chương I	23
CHƯƠNG II: GIỚI THIỆU VÀ NGHIÊN CỨU CNN	24
2.1 Khát quát bài toán nhận diện hình ảnh và CNN	24
2.1.1 Bài toán nhận diện hình ảnh	24
2.1.2 Học máy, học sâu và CNN	25
2.2 Một số nghiên cứu có liên quan	36
2.2.1 Quá trình phát triển của các kiến trúc CNN	36
2.2.2. Các mạng CNN tiêu biểu	37
2.3 Học chuyển giao	41
2.4 Mô hình kiến trúc ResNet và áp dụng vào bài toán phân loại ảnh	42
2.4.1. Giới thiệu về ResNet	42
2.4.2. Batch Normalization	43
2.4.3 Kết nối tắt (Skip Connection)	43
2.4.4. Mô hình ResNet50 áp dụng vào bài toán	45
2.5 Kết luận chương II	48
CHƯƠNG III: THỰC NGHIỆM, ĐÁNH GIÁ VÀ XÂY DỰNG HỆ THỐNG NHẬN DẠNG HÌNH ẢNH TRONG THƯƠNG MẠI ĐIỆN TỬ	49
3.1 Dữ liệu thử nghiệm	49
3.2 Tiến hành thực nghiệm và kết quả	50
3.2.1 Tiến hành thực nghiệm	50
3.2.2. Kết quả thực nghiệm	53
3.3 Đánh giá kết quả	58
3.3.1. Đánh giá bằng độ Accuracy	58
3.3.2. Đánh giá bằng Confusion matrix	59
3.2.2. Đánh giá bằng Precision, Recall, F1-Score	60
3.4 Áp dụng xây dựng hệ thống	63

3.4.1 Xây dựng web nhận dạng sản phẩm	63
3.4.2 Tạo web với Plask.....	63
3.4.3 Trang web phân loại sản phẩm thương mại điện tử	64
3.4.4 Trang web nhận dạng sản phẩm thương mại điện tử.....	66
3.5 Kết luận chương III	68
KẾT LUẬN	69
1. Kết quả đạt được:	69
2. Hướng nghiên cứu tiếp theo:	69

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Viết đầy đủ	Giải nghĩa
API	Application Programming Interface	Giao diện lập trình ứng dụng
B2A	Business To Goverment	Doanh nghiệp với Chính phủ
B2B	Business To Business	Doanh nghiệp với doanh nghiệp
B2C	Business To Consumer	Doanh nghiệp với Khách hàng
C2A	Consumer To Goverment	Khách hàng với Chính Phủ
C2B	Consumer To Business	Khách hàng với Doanh nghiệp
C2C	Consumer To Consumer	Khách hàng với Khách hàng
CNN	Convolutional Neural Network	Mạng nơ ron tích chập
ConvNet	Convolutional Neural Network	Mạng nơ ron tích chập
DL	Deep Learning	Học sâu
E – Commerce	Electronic commerce	Thương mại điện tử
EDI	Electronic Data Interchange	Trao đổi dữ liệu điện tử
ELU	Exponential Linear Unit	Đơn vị tuyến tính lũy thừa
FC	Fully connected	Kết nối đầy đủ
HTTPS	Hypertext Transfer Protocol Secure	Giao thức truyền siêu văn bản an toàn
IDE	Integrated Development Environment	Môi trường phát triển tích hợp
KNN	K-nearest Neighbor	K láng giềng
ML	Machine Learning	Học máy
NN	Neural network	Mạng nơ ron nhân tạo
ReLU	Rectified Linear Unit	Đơn vị tuyến tính sửa chữa

SSL	Secure Sockets Layer	Bảo mật tầng giao vận
SVM	Support vector machine	Máy vectơ hỗ trợ
TMĐT	Thương mại điện tử	Thương mại điện tử
VTO	World Trade Organization	Tổ chức thương mại thế giới
WSGI	Web Server Gateway Interface	Giao diện cổng máy chủ Web

DANH MỤC BẢNG BIỂU/HÌNH VẼ

Bảng 3.1. Chi tiết tập dữ liệu thực nghiệm	49
Hình 1.1. Đặc trưng của Thương mại điện tử	8
Hình 1.2. Bốn hình thức thanh toán điện tử phổ biến tại Việt Nam	20
Hình 2.1. Minh họa hệ thống phân loại chó và mèo	24
Hình 2.2. Sơ đồ phân nhóm thuật toán Machine learning	26
Hình 2.3. Mô hình hồi quy Logic.....	28
Hình 2.4 Mô hình mạng neural network	29
Hình 2.5. Mô tả lấy chập dùng bộ lọc kích thước 5x5.....	31
Hình 2.6. Mô tả chi tiết lấy chập dùng bộ lọc kích thước 5x5.....	32
Hình 2.7. Mô tả bước lấy chập của mạng nơron dùng 3 bộ lọc kích thước 5x5	32
Hình 2.8. Ví dụ sử dụng max-pooling.....	34
Hình 2.9. Mô tả cách thực hiện max-pooling với padding	35
Hình 2.10. Mô hình mạng neural tích chập.....	36
Hình 2.11. Kiến trúc LeNet.....	38
Hình 2.12. Hàm ReLu	39
Hình 2.13. Minh họa phương pháp dropout.....	39
Hình 2.14. Kiến trúc mạng AlexNet	40
Hình 2.15. Kiến trúc VGG-16.....	41
Hình 2.16. Khối ResNet thông thường và khối ResNet với tầng tích chập 1x1	44
Hình 2.17. Khối xác định (Identity block)	44
Hình 2.18. Sự khác biệt giữa một khối thông thường (trái) và một khối xác định(phải)	45
Hình 2.19. Kết nối tắt qua 3 lớp của RestNet50 so với 2 lớp của RestNet34.....	46
Hình 2.20. Kiến trúc tóm tắt của mạng ResNet50.	46
Hình 2.21. Kiến trúc chi tiết của RestNet50	46
Hình 3.1. Cây thư mục của tập dữ liệu thực nghiệm	50

Hình 3.2. Các ảnh của mặt hàng "Cháo" trong tập dữ liệu	50
Hình 3.3. Tổng quan mô hình ResNet50 được sử dụng.....	51
Hình 3.4. Model Checkpoint.....	52
Hình 3.5. Các epochs trong quá trình đào tạo mô hình.....	53
Hình 3.6. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 1.....	54
Hình 3.7. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 2.....	55
Hình 3.8. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 3.....	55
Hình 3.9. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 1	56
Hình 3.10. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 2....	56
Hình 3.11. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 3....	57
Hình 3.12. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm thời trang	57
Hình 3.13. Độ chính xác Accuracy trên tập kiểm thử.....	58
Hình 3.14. Biểu đồ đường độ chính xác accuracy qua các epochs	59
Hình 3.15. Confusion matrix của mô hình trên tập test	60
Hình 3.16. Precision, Recall, F1-Score của mô hình trên tập test	62
Hình 3.17. Cách Flask Framework hoạt động	64
Hình 3.18. File giao diện index.html của Web nhận dạng sản phẩm TMĐT	65
Hình 3.19. API RESTful GET và POST của Flask trong Web nhận dạng sản phẩm thương mại điện tử	65
Hình 3.20. Trang giao diện web nhận dạng sản phẩm TMĐT	66
Hình 3.21. Cửa sổ chọn file ảnh truy vấn trên thư mục máy tính.....	67
Hình 3.22. Web hiển thị ảnh sản phẩm và kết quả nhận dạng.....	67
Hình 3.23. Một kết quả nhận dạng sản phẩm khác của web.....	68

MỞ ĐẦU

1. Lý do chọn 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,...

Mỗi người, khi có một nhu cầu mua bán hoặc giải trí, để thực hiện người đó có thể đến một địa điểm bán hàng hoặc vui chơi, nơi đó có những nhân viên có thể tư vấn về vấn đề của khách hàng hoặc khách hàng có thể thỏa thích xem qua những sản phẩm trên kệ hàng. Hoặc người có nhu cầu sử dụng internet để tìm kiếm.

Người thích đọc sách, các trang web đọc sách Online, mua bán sách sẽ là địa chỉ truy cập thường xuyên. Người dùng có thể tìm kiếm ra cuốn sách mình muốn đọc nhưng có thể, có nhưng cuốn sách hay phù mà họ không biết đến.

Tương tự như vậy, người thích mua sắm có thể tìm kiếm những gì mình cần mua trên các trang bán hàng, nhưng có thể có những vật dụng khác mà nhất thời chưa nghĩ ra hoặc chưa được biết đến phù hợp với họ.

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. Tổng quan về vấn đề nghiên cứu

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”.

Trong bài báo của Hiranmay Ghosh, tác giả có trình bày một cách tiếp cận dựa trên ví dụ tìm kiếm một sản phẩm cụ thể dựa trên hình ảnh tìm ra tính chất của sản phẩm đó. Người dùng có thể chụp ảnh gói sản phẩm bằng điện thoại di động hoặc webcam và gửi đến các sàn thương mại điện tử để tìm thông tin chi tiết về sản phẩm. Sau đó hệ thống sẽ tìm kiếm cơ sở dữ liệu hình ảnh sản phẩm cho các tính năng trực quan đặc biệt trên hình ảnh truy vấn để xác định sản phẩm mà người dùng mong muốn. Từ đó giúp cho người dùng thuận tiện, nhanh chóng hơn so với các cách truyền thống trong việc tìm kiếm trên sàn thương mại điện tử.

Hiện nay, Amazon Rekognition Custom Labels đã giúp người dùng có thể xác định các đối tượng và cảnh trong hình ảnh cụ thể cho nhu cầu mua hàng của người dùng. Gần đây, Amazon đã được mở rộng chức năng nhận dạng hình ảnh sản phẩm. Người dùng có thể tìm thấy sản phẩm được trưng bày trong các bài đăng trên mạng xã hội và sàn thương mại điện tử. Amazon đã phân loại riêng biệt chi tiết như các bộ phận máy móc trong dây chuyền lắp ráp hoặc phát hiện các sản phẩm trong hình ảnh.

3. Mục đích nghiên cứu

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.

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. Đối tượng và phạm vi nghiên cứu

Đề tài nghiên cứu lý thuyết và ứng dụng trong các sàn thương mại điện tử tại Cục Thương mại điện tử và Kinh tế số (Bộ Công Thương).

5. Phương pháp nghiên cứu

-Phương pháp 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.

-Phương pháp 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.

Từ mục tiêu, nhiệm vụ nghiên cứu, luận văn sẽ được cấu trúc với ba chương nội dung chính như sau:

Chương 1: Tổng quan về thương mại điện tử Việt Nam và Thế giới

Chương 2: Giới thiệu và nghiên cứu CNN cho bài toán nhận diện hình ảnh

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

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. Năm 1979, Viện Tiêu chuẩn Quốc gia Hoa Kỳ đã phát triển ASC X12 như một tiêu chuẩn chung cho các doanh nghiệp để chia sẻ tài liệu thông qua mạng điện tử.

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

Thương mại điện tử là việc mua và bán hàng hóa và dịch vụ, hoặc chuyển tiền hoặc dữ liệu, qua một mạng điện tử, chủ yếu là internet. Các giao dịch kinh doanh này xảy ra dưới dạng doanh nghiệp với doanh nghiệp (B2B), doanh nghiệp với người tiêu dùng (B2C), người tiêu dùng với người tiêu dùng hoặc người tiêu dùng với doanh nghiệp. Các thuật ngữ thương mại điện tử và kinh doanh điện tử thường được sử dụng thay thế cho nhau.

Hiện nay, khái niệm thương mại điện tử của WTO là chính xác nhất. “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”. Việc mua bán hàng hóa trên Shopee, Lazada, Tiki hoặc qua website thương mại là các ví dụ về thương mại điện tử nổi bật.

Các hoạt động thương mại điện tử chủ yếu bao gồm:

- Mua bán và trao đổi hàng hóa, dịch vụ trực tuyến
- Mua bán vé trực tuyến
- Thanh toán online
- Chăm sóc và hỗ trợ khách hàng online

1.1.3 Cách thức hoạt động

Thương mại điện tử được cung cấp bởi internet, nơi khách hàng có thể truy cập vào một cửa hàng trực tuyến để duyệt qua và đặt hàng cho các sản phẩm hoặc dịch vụ thông qua thiết bị của riêng họ.

Khi đơn đặt hàng được đặt, trình duyệt web của khách hàng sẽ giao tiếp qua lại với máy chủ lưu trữ trang web của cửa hàng trực tuyến. Dữ liệu liên quan đến đơn đặt hàng sau đó sẽ được chuyển tiếp đến một máy tính trung tâm được gọi là trình quản lý đơn hàng- sau đó được chuyển tiếp đến cơ sở dữ liệu quản lý mức tồn kho, hệ thống người bán quản lý thông tin thanh toán (sử dụng các ứng dụng như PayPal) và máy tính ngân hàng - trước khi quay trở lại người quản lý đơn đặt hàng. Điều này nhằm đảm bảo rằng lượng hàng tồn kho trong cửa hàng và tiền của khách hàng có đủ để xử lý đơn đặt hàng. 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. Tại thời điểm này, các sản phẩm hữu hình và / hoặc kỹ thuật số có thể được chuyển đến khách hàng, hoặc quyền truy cập vào một dịch vụ có thể được cấp.

Các nền tảng lưu trữ các giao dịch thương mại điện tử có thể bao gồm các thị trường trực tuyến mà người bán chỉ cần đăng ký, chẳng hạn như Amazon.com; phần mềm như một công cụ dịch vụ (SaaS) cho phép khách hàng 'thuê' cơ sở hạ tầng của

hàng trực tuyến; hoặc các công cụ mã nguồn mở để các công ty sử dụng phát triển nội bộ để quản lý.

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).

- **Doanh nghiệp với Doanh nghiệp (B2B)**

Đây là hình thức thương mại điện tử đề cập đến việc trao đổi điện tử các sản phẩm, dịch vụ hoặc thông tin giữa các doanh nghiệp chứ không phải giữa doanh nghiệp và người tiêu dùng. Ví dụ bao gồm các thư mục trực tuyến và các trang web trao đổi sản phẩm và nguồn cung ứng cho phép các doanh nghiệp tìm kiếm các sản phẩm, dịch vụ và thông tin cũng như bắt đầu giao dịch thông qua các giao diện mua sắm điện tử.

Vào năm 2017, Forrester Research dự đoán rằng thị trường thương mại điện tử B2B sẽ đạt 1,1 nghìn tỷ đô la Mỹ vào năm 2021, chiếm 13% tổng doanh số bán hàng B2B trên toàn quốc.

- **Doanh nghiệp với Khách hàng (B2C)**

Hình thức này là một phần bán lẻ của thương mại điện tử trên internet. Đó là khi doanh nghiệp bán sản phẩm, dịch vụ hoặc thông tin trực tiếp cho người tiêu dùng. Thuật ngữ này phổ biến trong thời kỳ bùng nổ dot-com vào cuối những năm 1990, khi các nhà bán lẻ và người bán hàng hóa trực tuyến là một điều mới lạ.

Ngày nay, có vô số cửa hàng và trung tâm thương mại ảo trên internet bán tất cả các loại hàng tiêu dùng. Ví dụ được công nhận nhiều nhất về các trang web này là Amazon, công ty thống trị thị trường B2C.

- **Khách hàng với Khách hàng (C2C)**

Đây là một loại hình thương mại điện tử trong đó người tiêu dùng trao đổi sản phẩm, dịch vụ và thông tin với nhau trực tuyến. Các giao dịch này thường được thực

hiện thông qua một bên thứ ba cung cấp nền tảng trực tuyến để thực hiện các giao dịch.

Đấu giá trực tuyến và quảng cáo đã phân loại là hai ví dụ về nền tảng C2C, với eBay và Craigslist là hai trong số các nền tảng này phổ biến nhất. Bởi vì eBay là một doanh nghiệp, hình thức thương mại điện tử này còn có thể được gọi là C2B2C - người tiêu dùng với người tiêu dùng.

- Khách hàng với Doanh nghiệp (C2B)

Một loại hình thương mại điện tử trong đó người tiêu dùng cung cấp sản phẩm và dịch vụ của họ trực tuyến để các công ty đặt giá thầu và mua. Điều này ngược lại với mô hình thương mại truyền thống của B2C.

Một ví dụ phổ biến về nền tảng C2B là một thị trường bán ảnh miễn phí bản quyền, hình ảnh, phương tiện truyền thông và các yếu tố thiết kế, chẳng hạn như iStock. Một ví dụ khác là bảng công việc.

- Doanh nghiệp với chính phủ (B2A)

Đề cập đến các giao dịch được thực hiện trực tuyến giữa các công ty và cơ quan hành chính công hoặc cơ quan chính phủ. Nhiều nhánh của chính phủ phụ thuộc vào các dịch vụ hoặc sản phẩm điện tử theo cách này hay cách khác, đặc biệt là khi liên quan đến các văn bản pháp lý, sổ đăng ký, an sinh xã hội, tài chính và việc làm. Các doanh nghiệp có thể cung cấp những thứ này dưới dạng điện tử. Các dịch vụ B2A đã phát triển đáng kể trong những năm gần đây khi đầu tư vào các khả năng của chính phủ điện tử.

- Khách hàng với Chính phủ (C2A)

Một hình thức phổ biến khác về các giao dịch được thực hiện trực tuyến giữa người tiêu dùng cá nhân và cơ quan hành chính công hoặc cơ quan chính phủ. Chính phủ hiếm khi mua các sản phẩm hoặc dịch vụ từ người dân, nhưng các cá nhân thường sử dụng các phương tiện điện tử trong các lĩnh vực sau:

- Giáo dục. Phổ biến thông tin, đào tạo từ xa / bài giảng trực tuyến, v.v.
- An ninh xã hội. Phân phối thông tin, thanh toán, v.v.
- Các loại thuế. khai thuế, thanh toán, v.v.

- Sức khỏe. Đặt hẹn, cung cấp thông tin về bệnh tật, thanh toán dịch vụ y tế, v.v.

1.1.5 Đặc trưng

Hình vẽ dưới đây gồm những đặc trưng của Thương mại điện tử.



Hình 1.1. Đặc trưng của Thương mại điện tử

Cụ thể Thương mại điện tử có các đặc trưng như sau:

- **Phổ biến**

Thương mại điện tử là phổ biến, có nghĩa là, nó luôn có sẵn ở mọi nơi. Nó giúp thị trường tự do không bị giới hạn trong một không gian vật lý và giúp bạn có thể mua sắm từ máy tính (chẳng hạn như máy tính để bàn, máy tính xách tay). Kết quả được gọi là không gian thị trường.

Đối với người tiêu dùng, sự phổ biến giúp giảm chi phí giao dịch để khám phá sản phẩm trên thị trường. Người tiêu dùng có thể thu thập bất kỳ thông tin nào bất cứ khi nào và bất cứ nơi nào họ muốn, bất kể vị trí của họ. Người mua không còn cần thiết phải tốn thời gian và tiền bạc để đi chợ nữa. Nói chung, nó tiết kiệm năng lượng nhận thức cần thiết để chuyển đổi trong không gian thị trường.

- **Phạm vi tiếp cận toàn cầu**

Công nghệ thương mại điện tử cho phép một doanh nghiệp dễ dàng tiếp cận qua các ranh giới địa lý trên trái đất một cách thuận tiện và hiệu quả hơn nhiều so với thương mại truyền thống. Trên toàn cầu, các công ty đang thu được lợi nhuận và kết

quả kinh doanh lớn hơn bằng cách mở rộng kinh doanh với các giải pháp thương mại điện tử. Do đó, quy mô thị trường tiềm năng cho các thương gia thương mại điện tử xấp xỉ quy mô dân số trực tuyến.

- **Tiêu chuẩn chung**

Tiêu chuẩn chung là tiêu chuẩn được chia sẻ bởi tất cả các quốc gia trên thế giới. Đây là các tiêu chuẩn kỹ thuật của Internet để tiến hành thương mại điện tử. Nó cung cấp cho tất cả khả năng kết nối ở cùng một "cấp độ" và nó cung cấp các ngoại tác mạng sẽ mang lại lợi ích cho tất cả mọi người. Các tiêu chuẩn kỹ thuật phổ biến giúp giảm chi phí đầu vào và chi phí tìm kiếm tối thiểu.

- **Tính tương tác**

Công nghệ thương mại điện tử cho phép giao tiếp hai chiều giữa khách hàng và người bán, giúp nó có tính tương tác. Nó chứng tỏ là một tính năng quan trọng của công nghệ thương mại điện tử so với các công nghệ thương mại truyền thống của thế kỷ 20.

- **Mật độ thông tin**

Mật độ thông tin có nghĩa là tổng số lượng và chất lượng của thông tin có sẵn trên Internet cho tất cả người mua và người bán trên thị trường. Internet làm tăng mật độ thông tin. Mật độ thông tin cung cấp thông tin chất lượng tốt hơn cho người tiêu dùng và thương gia. Công nghệ thương mại điện tử làm tăng tính chính xác và kịp thời của thông tin. Ví dụ, cửa hàng flipkart.com có nhiều loại sản phẩm với giá cả.

- **Sự phong phú**

Sự phong phú đề cập đến mức độ phức tạp và nội dung của một thông điệp. Sự phong phú có nghĩa là tất cả các hoạt động thương mại và trải nghiệm, được thực hiện thông qua nhiều loại thông điệp. Ví dụ: văn bản, hình ảnh, video, âm thanh, liên kết, SMS (Dịch vụ tin nhắn ngắn), v.v.

- **Cá nhân hóa**

Công nghệ thương mại điện tử mang đến sự cá nhân hóa. Cá nhân hóa nghĩa là thiết kế thông điệp tiếp thị theo từng cá nhân cụ thể bằng cách tùy chỉnh thông điệp đó theo thông tin cá nhân của khách hàng như tên, sở thích và hồ sơ mua hàng trước

đây. Sản phẩm hoặc dịch vụ có thể được sửa đổi hoặc thay đổi theo sự lựa chọn của người dùng hoặc hồ sơ mua trước đây.

1.1.6 Ưu và nhược điểm của thương mại điện tử

❖ Ưu điểm:

Lợi ích của thương mại điện tử bao gồm tính khả dụng 24/24, tốc độ truy cập, sự sẵn có rộng rãi của hàng hóa và dịch vụ cho người tiêu dùng, khả năng tiếp cận dễ dàng và phạm vi tiếp cận quốc tế.

- Khả dụng. Ngoài việc ngừng hoạt động hoặc bảo trì theo lịch trình, các trang thương mại điện tử có sẵn 24/7, cho phép khách truy cập duyệt và mua sắm bất cứ lúc nào. Các doanh nghiệp truyền thống có xu hướng mở cửa trong một số giờ cố định và thậm chí có thể đóng cửa hoàn toàn vào những ngày nhất định.
- Tốc độ truy cập. Mặc dù những người mua sắm trong một cửa hàng thực có thể bị chậm lại bởi đám đông, nhưng các trang web thương mại điện tử chạy nhanh chóng, điều này được xác định bằng việc cân nhắc tính toán và băng thông trên cả thiết bị tiêu dùng và trang web thương mại điện tử. Trang sản phẩm và trang giỏ hàng tải trong vài giây hoặc ít hơn. Một giao dịch thương mại điện tử có thể bao gồm một vài cú nhấp chuột và mất chưa đầy năm phút.
- Tính khả dụng rộng rãi. Khẩu hiệu đầu tiên của Amazon là "Hiệu sách lớn nhất Trái đất." Họ có thể đưa ra yêu cầu này bởi vì họ là một trang web thương mại điện tử chứ không phải một cửa hàng thực phải để từng cuốn sách trên kệ của nó. Thương mại điện tử cho phép các thương hiệu cung cấp nhiều loại sản phẩm, sau đó được vận chuyển từ kho hàng sau khi mua hàng. Khách hàng có thể thành công hơn trong việc tìm kiếm những gì họ muốn.
- Khả năng tiếp cận dễ dàng. Khách hàng mua sắm tại một cửa hàng thực có thể gặp khó khăn trong việc xác định lối đi nào của một sản phẩm cụ thể. Trong thương mại điện tử, khách truy cập có thể duyệt qua các trang danh mục sản phẩm và sử dụng tính năng tìm kiếm trên trang web để tìm sản phẩm ngay lập tức.

- Tiếp cận quốc tế. Các doanh nghiệp truyền thống bán hàng cho những khách hàng ghé thăm cửa hàng của họ. Với thương mại điện tử, doanh nghiệp có thể bán cho bất kỳ khách hàng nào có thể truy cập web. Thương mại điện tử có tiềm năng mở rộng cơ sở khách hàng của doanh nghiệp
- Chi phí thấp hơn. Các doanh nghiệp thương mại điện tử thường tránh được chi phí liên quan đến các cửa hàng thực, chẳng hạn như tiền thuê, hàng tồn kho và nhân viên thu ngân, mặc dù họ có thể phải chịu chi phí vận chuyển và kho hàng
- Cá nhân hóa và đề xuất sản phẩm. Các trang thương mại điện tử có thể theo dõi lịch sử duyệt, tìm kiếm và mua hàng của khách. Họ có thể sử dụng dữ liệu này để đưa ra các đề xuất sản phẩm hữu ích và được cá nhân hóa, đồng thời có được thông tin chi tiết có giá trị về thị trường mục tiêu. Ví dụ bao gồm các phần của trang sản phẩm Amazon có nhãn "Thường được mua cùng nhau" và "Khách hàng đã xem mặt hàng này cũng đã xem".

❖ **Nhược điểm:**

- Dịch vụ khách hàng hạn chế. Nếu khách hàng có thắc mắc hoặc vấn đề trong cửa hàng thực, họ có thể gặp nhân viên bán hàng, thu ngân hoặc quản lý cửa hàng để được trợ giúp. Trong cửa hàng thương mại điện tử, dịch vụ khách hàng có thể bị hạn chế: Trang web chỉ có thể cung cấp hỗ trợ vào những giờ nhất định trong ngày hoặc cuộc gọi đến số điện thoại dịch vụ khách hàng có thể khiến khách hàng phải chờ đợi.
- Không thể chạm hoặc nhìn. Mặc dù hình ảnh trên trang web có thể mang lại cảm nhận tốt về sản phẩm, nhưng nó khác với việc trải nghiệm sản phẩm đó một cách "trực tiếp", chẳng hạn như phát nhạc trên loa, đánh giá chất lượng hình ảnh của tivi hoặc mặc thử áo sơ mi hoặc váy. Thương mại điện tử có thể khiến người tiêu dùng nhận được các sản phẩm khác với mong đợi của họ, dẫn đến lợi nhuận. Trong một số trường hợp, khách hàng phải chịu gánh nặng về chi phí vận chuyển mặt hàng bị trả lại cho nhà bán lẻ.

- Thời gian chờ đợi. Nếu khách hàng nhìn thấy một món đồ mà họ thích trong cửa hàng, khách hàng sẽ trả tiền cho món đồ đó và sau đó sẽ về nhà với món đồ đó. Với thương mại điện tử, có một thời gian chờ đợi để sản phẩm được chuyển đến địa chỉ của khách hàng. Mặc dù thời hạn giao hàng đang giảm vì việc giao hàng vào ngày hôm sau hiện khá phổ biến, nhưng nó không phải là tức thời.

1.1.7 Lợi ích thương mại điện tử đến doanh nghiệp

- **Rào cản nhập cảnh thấp**

Trong thế giới ngày nay, các công ty lớn và nhỏ đều có cơ hội thành lập và tiến hành kinh doanh trên Internet. Chi phí truy cập Internet của các công ty là rất nhỏ (Rất nhỏ) vì họ không cần mặt bằng cho thuê. Tất cả các hoạt động kinh doanh trên Internet là ảo có nghĩa là không cần số lượng lớn nhân viên để tiến hành kinh doanh.

- **Tăng Thị phần Tiềm năng:**

Các doanh nghiệp đang tăng thị phần của họ bằng cách kích hoạt internet kinh doanh của họ. Các doanh nghiệp trực tuyến được tiếp cận bất kỳ lúc nào với thị trường quốc tế.

- **Quảng cáo với chi phí thấp**

Internet cung cấp quảng cáo với chi phí thấp hơn so với quảng cáo trên báo chí hoặc truyền hình. Trong thế giới ngày nay, Internet đã trở thành phương tiện quảng cáo rẻ tiền được các công ty sử dụng cho thương mại. Các phương pháp quảng cáo khác nhau là: e-mail, biểu ngữ, cửa sổ bật lên, video và âm thanh hấp dẫn, v.v.

- **Lợi ích chiến lược:**

Kinh doanh hỗ trợ thương mại điện tử có nhiều lợi ích chiến lược vì chúng:

- Giảm chi phí chuẩn bị thư, chuẩn bị tài liệu và nhập dữ liệu.
- Dễ dàng tìm ra lỗi.
- Giảm chi phí gọi qua điện thoại.
- Giảm thời gian giao hàng và chi phí lao động.
- Giảm chi phí quản lý và nhập dữ liệu.

- **Phạm vi tiếp cận toàn cầu:**

Doanh nghiệp hỗ trợ thương mại điện tử có khả năng tiếp cận toàn cầu với chi phí thấp. Họ có thể gửi tin nhắn trên toàn thế giới bất cứ lúc nào. Vì các doanh nghiệp trực tuyến được tiếp cận trên toàn cầu nên thương mại điện tử giúp thu hút người tiêu dùng và khách hàng doanh nghiệp mới từ mọi nơi trên thế giới.

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

1.2.1 Thương mại điện tử toàn cầu.

Trong thập kỷ qua, việc sử dụng rộng rãi các nền tảng thương mại điện tử như Amazon và eBay đã góp phần vào sự tăng trưởng đáng kể trong lĩnh vực bán lẻ trực tuyến. 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

Đầu tiên thương mại điện tử toàn cầu là việc bán các sản phẩm hoặc dịch vụ xuyên biên giới địa chính trị từ quốc gia xuất xứ của một công ty, thường được định nghĩa là vị trí thành lập hoặc kết hợp của công ty. Sản phẩm hoặc dịch vụ được bán vào các thị trường không phải bản địa thông qua bán hàng và tiếp thị trực tuyến.

Dữ liệu tích lũy dự đoán doanh số thương mại điện tử trên toàn thế giới tăng 16,8% trong khoảng thời gian được theo dõi gần đây nhất.

Những lợi thế của thương mại điện tử quốc tế là:

- Mở rộng ra thị trường nước ngoài dễ dàng hơn
- Sản phẩm phù hợp với thị trường dễ tìm hơn
- Chu kỳ bán hàng B2B ngắn hơn
- Xây dựng nhanh hơn sự hiện diện quốc tế
- Giảm rào cản gia nhập

Như Harvard Business Review đã viết: "Các nhà lãnh đạo doanh nghiệp đang cố gắng thích nghi với một thế giới mà ít người có thể tưởng tượng được chỉ một năm trước. Huyền thoại về một thế giới không biên giới đã sụp đổ. Các trụ cột truyền thống của thị trường mở - Hoa Kỳ và Anh - đang lung lay, và Trung Quốc đang tự định vị mình là người bảo vệ vững chắc nhất của toàn cầu hóa. " 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.

Hai năm trước, chỉ có 17,8% doanh số được thực hiện từ mua hàng trực tuyến. Con số này dự kiến sẽ đạt 21% vào năm 2022, tăng 17,9% thị phần thương mại điện tử trong vòng hai năm. Tăng trưởng dự kiến sẽ tiếp tục, đạt 24,5% vào năm 2025, tức là tăng 6,7 điểm phần trăm chỉ trong vòng 5 năm.

- **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. Xu hướng tiếp tục:

- Mỹ Latinh đạt 85 tỷ USD doanh thu từ thương mại điện tử vào năm 2021, tăng 25% so với 68 tỷ USD vào năm 2020.
- Thị trường thương mại điện tử Ấn Độ dự kiến sẽ tăng lên 111,4 tỷ đô la vào năm 2025, tăng từ 46,2 tỷ đô la vào năm 2020.
- Nga, Anh và Philippines đã chứng kiến mức tăng trưởng doanh số thương mại điện tử hơn 20% vào năm 2021.
- Trung Quốc tiếp tục dẫn đầu thị trường thương mại điện tử toàn cầu, chiếm 52,1% tổng doanh số thương mại điện tử bán lẻ trên toàn thế giới, với tổng doanh số bán hàng trực tuyến chỉ hơn 2 nghìn tỷ đô la vào năm 2021. Nước này cũng có nhiều người mua kỹ thuật số nhất thế giới, 824,5 triệu người, chiếm 38,5% tổng số toàn cầu.
- Thị trường thương mại điện tử của Mỹ được dự báo sẽ đạt hơn 875 tỷ USD vào năm 2022, hơn một phần ba so với thị trường của Trung Quốc. Sau Trung Quốc và Mỹ, thị trường thương mại điện tử lớn thứ ba là Vương quốc Anh, chiếm

4,8% thị phần thương mại điện tử bán lẻ. Tiếp theo là Vương quốc Anh (3%) và Hàn Quốc (2,5%).

Năm thị trường thương mại điện tử hàng đầu không thay đổi kể từ năm 2018. Xu hướng từ eMarketer cho thấy những thị trường này sẽ ở trong top năm cho đến năm 2025.

Casey Armstrong, CMO tại thương hiệu thực hiện thương mại điện tử ShipBob, cho biết thêm, “Mặc dù tập trung nhiều vào các trung tâm thương mại điện tử trên khắp Hoa Kỳ và Canada, nhưng có rất nhiều điều để học hỏi từ những người chơi quốc tế lớn khác, những người đang chứng kiến tốc độ phát triển thậm chí còn nhanh hơn trong thương mại điện tử ”

Ông cho biết thêm, “Người bán có thể thay đổi nơi họ bán dựa trên dữ liệu này và nhu cầu về thương mại điện tử từ các quốc gia này. Tại ShipBob, đó là lý do tại sao chúng tôi đã mở các trung tâm thực hiện ở Canada và Vương quốc Anh và sắp mở một trung tâm khác ở Úc”

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

Là một trong những nền kinh tế Internet phát triển nhanh nhất trong khu vực Đông Nam Á, giá trị thị trường thương mại điện tử của Việt Nam được dự báo sẽ đạt 39 tỷ đô la Mỹ vào năm 2025, đứng thứ hai chỉ sau Indonesia. Tại Việt Nam, dân số kỹ thuật số hiện nay và mức độ thâm nhập internet ngày càng tăng tạo điều kiện thích hợp cho các doanh nghiệp thương mại điện tử phát triển mạnh mẽ. Do đó, tỷ trọng thương mại điện tử trong tổng doanh thu bán lẻ trong nước đang tăng nhanh trong khi các kênh trực tuyến đã vượt qua các kênh bán lẻ hiện đại về mức tăng trưởng trong thị trường FMCG.

- **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 chút so với các thị trường phát triển hơn trong báo cáo của chúng tôi. Tuy nhiên, nó tự hào có một trong những dự đoán tăng trưởng cao hơn trong nghiên cứu của chúng tôi, khi một quốc gia gồm những người tiêu dùng trẻ tuổi ngày càng chuyển sang các phương pháp kỹ thuật số để mua sắm.

Hiện tại, thương mại điện tử chiếm 6% tổng doanh số bán lẻ ở Việt Nam và 60% dân số vẫn chưa thực hiện giao dịch mua hàng trực tuyến đầu tiên. Tuy nhiên, việc áp dụng thương mại điện tử ngày càng tăng được hỗ trợ bởi một thế hệ sắp tới của những người bản địa kỹ thuật số bắt đầu kiếm tiền và tiêu tiền — 37,9% dân số từ 24 tuổi trở xuống.

Mạng xã hội là một cách chính để tiếp cận nhóm người tiêu dùng Việt Nam tiếp theo này, với 66% người mua sắm trực tuyến đã sử dụng Facebook để mua các mặt hàng. 5 Với dân số 97,4 triệu người, ước tính cả nước sẽ có khoảng 50,9 triệu người dùng mạng xã hội vào năm 2021, so với 43,8 triệu năm 2017. Đối với công dân từ 16 đến 64 tuổi, 90% sử dụng Facebook, với Ứng dụng trò chuyện YouTube và Zalo trong nước cũng rất phổ biến.

Các thị trường điện tử quốc tế như Shopee của Singapore và Lazada thuộc sở hữu của Alibaba hiện đang giữ vị trí hàng đầu trong thị trường thương mại điện tử Việt Nam. Trong khi đó, ngày càng có nhiều trang thương mại điện tử Việt Nam đăng ký mới, đặc biệt là hình thức chợ điện tử. Thương hiệu địa phương Thế Giới Di Động nổi lên như một trong những trang thương mại điện tử phổ biến nhất trong những năm gần đây, chuyên về điện tử tiêu dùng và thiết bị CNTT. Nhìn chung, doanh thu giữa doanh nghiệp với người tiêu dùng (B2C) của lĩnh vực thương mại điện tử đã tăng từ dưới một tỷ đô la Mỹ vào năm 2012 lên khoảng 12 tỷ đô la Mỹ vào năm 2020.

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

Lĩnh vực du lịch, di chuyển và lưu trú từng chiếm chi tiêu cho thương mại điện tử lớn nhất trong số những người sử dụng Internet tại Việt Nam trước đại dịch COVID-19. Tuy nhiên, lĩnh vực này đã ghi nhận mức giảm 45% trong chi tiêu tiêu dùng vào năm 2020 do tác động của đại dịch COVID-19 đối với ngành du lịch và lữ hành. Đồng thời, các danh mục thương mại điện tử hàng tiêu dùng khác có mức tăng trưởng đáng kể, với các sản phẩm thực phẩm và đồ uống có mức tăng chi tiêu tiêu dùng cao nhất. Trong thời kỳ đại dịch, ngày càng nhiều người tiêu dùng Việt Nam sẵn sàng mua sắm trực tuyến các nhu yếu phẩm và sản phẩm tươi sống. Kết quả là,

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**

So với các nước láng giềng Đông Nam Á, Việt Nam là một quốc gia phụ thuộc nhiều vào tiền mặt. Khi nói đến mua sắm trực tuyến, tiền mặt khi nhận hàng vẫn là phương thức thanh toán phổ biến nhất của người Việt Nam. Mặc dù mức độ phổ biến của phương thức này đã giảm nhẹ, nhưng nó vẫn vượt qua thanh toán không dùng tiền mặt về mức độ sử dụng vào năm 2020 vì nó giảm thiểu rủi ro cho người mua sắm trực tuyến. 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. Với sự phát triển không ngừng của lĩnh vực fintech tại Việt Nam, thanh toán không dùng tiền mặt đang dần trở thành phương thức thanh toán phổ biến cho nhiều người dùng thương mại điện tử hơn nữa trong tương lai.

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

1.3.1 Cơ sở hạ tầng pháp lý

Thương mại điện tử chịu sự điều chỉnh của Nghị định 52/2013/NĐ-CP ban hành năm 2013. Bộ Công Thương (MOIT) là cơ quan quản lý các hoạt động Thương mại điện tử. Vào tháng 9 năm 2020, Bộ Công Thương đã ban hành Nghị định 85/2021/NĐ-CP sửa đổi bổ sung Nghị định 52. Tại Nghị định 85 có một số thay đổi đáng kể được đề xuất, bao gồm một về hoạt động Thương mại điện tử trên cơ sở cung cấp xuyên biên giới. 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, và
- (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.

Theo Nghị định 52/2013 / NĐ-CP có hiệu lực từ ngày 01 tháng 7 năm 2013, không được lợi dụng thương mại điện tử để kinh doanh hàng giả, buôn bán hàng hóa, cung cấp dịch vụ xâm phạm quyền sở hữu trí tuệ, kinh doanh hàng hóa, cung ứng dịch vụ trong danh mục hàng hóa, dịch vụ cấm kinh doanh. Tên miền cũng được bảo vệ theo luật này. Thanh tra Bộ Công Thương, Cục Quản lý thị trường, Thanh tra Sở Công Thương và các cơ quan nhà nước khác có quyền xử phạt vi phạm hành chính trong lĩnh vực thương mại điện tử theo trách nhiệm của mình quy định tại Luật Xử lý vi phạm hành chính và tài liệu liên quan.

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. Đến năm 2020 hạ tầng viễn thông quốc gia đã phủ rộng khắp toàn quốc, kể cả ở những vùng sâu, vùng xa, biên giới, hải đảo với hơn 800.000 km cáp quang và các trạm thu phát sóng. Chất lượng hạ tầng công nghệ thông tin, viễn thông ngày càng được nâng cao. Hệ thống băng thông rộng được phát triển mạnh. Chỉ số Internet toàn diện Việt Nam năm 2018 đứng thứ 43 trong tổng số 86 quốc gia, thứ hạng cao hơn một số nước trong khu vực (Indonexia đứng thứ 49, Philippin đứng thứ 54); tốc độ tải trung bình năm 2018 đạt trên 6,9Mbps, đứng thứ 75 trên tổng số 200 quốc gia được xếp hạng (cao hơn Indonexia ở mức 5,8 Mbps, Philipin ở mức 5,2 Mbps). Số thuê bao băng rộng cố định đạt hơn 13,58 triệu, trong đó hơn 12 triệu thuê bao sử dụng cáp quang FTTx, tốc độ truy nhập hơn 10Mbps. Tổng băng thông quốc tế đạt hơn 8,1Tbps với 06 tuyến cáp quang biển kết nối đi quốc tế. Giá cước dịch vụ Internet Việt Nam ở mức vừa

phải, cước dịch vụ internet băng thông rộng cố định tại Việt Nam ở mức thấp nhất trong khu vực châu Á Thái Bình Dương (quy đổi theo sức mua tương đương)

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. Mạng di động 5G cũng đã được cấp phép thử nghiệm và thành công bước đầu. Số lượng các thuê bao tăng với tốc độ rất cao 30-40% năm, đến nay đã có hơn 100 triệu thuê bao điện thoại cố định và di động, hàng chục triệu thuê bao Internet. 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

Việt Nam là thị trường tiềm năng do có mức độ phủ sóng internet cao trong khi chi phí sử dụng internet thấp, đồng thời tỷ lệ người dân sử dụng các thiết bị thông minh tăng nhanh chóng.

Kể từ khi dịch bệnh bùng phát, thói quen mua sắm và thanh toán của người tiêu dùng Việt Nam đã thay đổi đáng kể khi người dân mong muốn có những trải nghiệm mua sắm an toàn, tiện lợi hơn, tránh các rủi ro trong giao dịch và mất mát tài sản. Theo thống kê của Visa năm 2020, người Việt Nam dành trung bình 3,1 giờ mỗi ngày để dùng các ứng dụng trực tuyến, nhưng trong thời gian xa cách xã hội, con số đó đã tăng vọt lên 4,2 giờ một ngày vào lúc cao điểm.

Hình thức thanh toán điện tử phổ biến tại Việt Nam:

- Thanh toán điện tử bằng thẻ ngân hàng

Hình thức thanh toán này là hình thức điện tử đầu tiên và đặc trưng nhất trong thị trường thanh toán điện tử. Tỷ lệ thanh toán bằng hình thức này hiện nay chiếm 90% trong tổng tỷ lệ thanh toán điện tử.

- Thanh toán qua cổng thanh toán

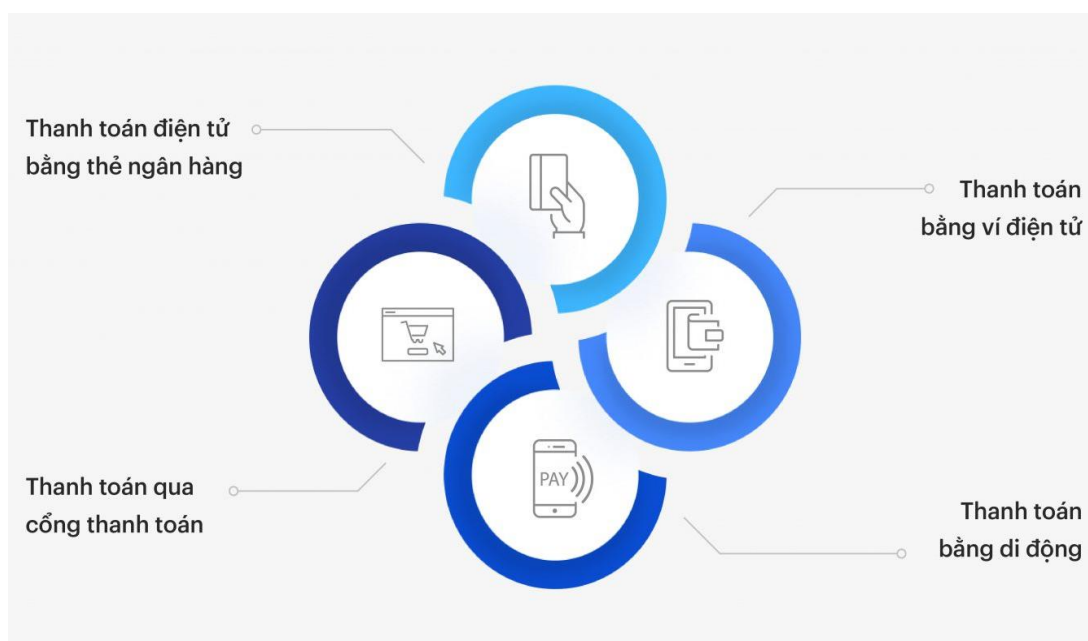
Hình thức thanh toán này phần lớn thường xuất hiện ở các trang thương mại điện tử. Người tiêu dùng chọn hình thức phù hợp và thao tác theo các bước hướng dẫn để hoàn thành giao dịch.

- Thanh toán bằng ví điện tử

Theo thống kê, 85% người tiêu dùng Việt Nam sở hữu ít nhất một ví điện tử hoặc ứng dụng thanh toán, trong đó hơn 42% người tiêu dùng sử dụng thanh toán thông qua thiết bị di động. Hầu hết các ví điện tử đều liên kết với tài khoản ngân hàng, các ứng dụng thương mại điện tử để mở rộng hệ sinh thái khách hàng và gia tăng trải nghiệm người dùng.

- Thanh toán bằng di động

Ngày 9/3/2021, Thủ tướng Chính phủ ký Quyết định 316/QĐ-TTg về phê duyệt triển khai thí điểm dùng tài khoản viễn thông (Mobile) để thanh toán cho hàng hóa dịch vụ. Hình thức Mobile Money hướng tới đối tượng khách hàng là người dân vùng sâu, vùng xa chưa có tài khoản ngân hàng hoặc hạ tầng internet phát triển. Mobile Money cung cấp dịch vụ thanh toán cho khách hàng thông qua tài khoản liên kết với số điện thoại của người dùng và được quản lý bởi công ty cung cấp dịch vụ viễn thông là Tập đoàn Viettel.



Hình 1.2. Bốn hình thức thanh toán điện tử phổ biến tại Việt Nam

1.3.4 Bảo mật trong thương mại điện tử

- Sử dụng HTTPS

Giao thức truyền siêu văn bản Bảo mật (HTTPS) là sự kết hợp của Giao thức truyền siêu văn bản (HTTP) với giao thức Lớp cổng bảo mật (SSL) / Bảo mật lớp truyền tải (TLS). TLS là một giao thức xác thực và bảo mật được triển khai rộng rãi trong các trình duyệt và máy chủ Web. SSL hoạt động bằng cách sử dụng khóa công khai để mã hóa dữ liệu được truyền qua kết nối SSL. Hầu hết các trình duyệt Web đều hỗ trợ SSL. Nó cho phép bạn giao tiếp an toàn với máy chủ web.

Đồng thời, chứng chỉ SSL của HTTPS còn có những ích lợi lớn đối với một website TMĐT:

- Hiện khóa xanh bảo mật, giúp tăng sự tin tưởng của người dùng.
- Có lợi cho SEO.
- Không bị trình duyệt (Chrome, Firefox, Safari) chặn vì thiếu bảo mật.

- Bảo mật server và Admin panel

Nếu sở hữu server riêng, việc áp dụng các biện pháp end-point security (bảo mật tại điểm cuối) là cần thiết cho mọi web TMĐT. Ngoài ra, các quản trị viên website cần có kiến thức để tự bảo vệ tài khoản Admin của mình bằng cách:

- Đặt mật khẩu khó, thay đổi mật khẩu định kỳ
- Cẩn trọng với những phần mềm, đường link, email không tin cậy.
- Phân quyền tài khoản rõ ràng.
- Đặt cảnh báo khi có địa chỉ IP lạ đăng nhập vào Admin panel.

- Bảo mật hệ thống thanh toán

Cổng thanh toán là một mục tiêu hấp dẫn của tin tặc, vì nó chứa thông tin giao dịch của người dùng và website TMĐT, đặc biệt là thông tin thẻ tín dụng. Để bảo mật cổng thanh toán, doanh nghiệp cần tuân thủ tối thiểu các chuẩn sau:

- Tích hợp chứng chỉ SSL mã hóa thông tin truyền tải
- Chuẩn bảo mật PCI DSS
- Mật khẩu OTP
- Mã hóa MD5 128 bit

-Cơ chế lưu token của người dùng

-Không lưu giữ thông tin thẻ của người dùng.

- Sử dụng phần mềm diệt virus và Malware (End-point security)

Sử dụng tường lửa – Web Các giải pháp bảo mật tại điểm cuối (end-point security) là cần thiết dành cho doanh nghiệp thương mại điện tử. Không chủ doanh nghiệp nào muốn công ty mình bị nhiễm virus, mã độc tống tiền, spyware hay trojan. Vậy nên mỗi máy tính làm việc cần trang bị một phần mềm diệt malware thật hiệu quả. Một vài phần mềm hiệu quả, đơn giản như Bitdefender, Windows Defender, Malwarebytes sẽ là ứng viên tiềm năng. Application Firewall

- Chứng chỉ SSL nâng cao (EV SSL)

Các website thương mại điện tử lớn nên sử dụng EV SSL thay vì chứng chỉ SSL thông thường. Mặc dù tính năng bảo mật của 2 loại chứng chỉ SSL này là như nhau. Nhưng việc sở hữu EV SSL sẽ giúp website TMĐT tránh việc bị giả mạo Phishing, từ đó bảo vệ người tiêu dùng tốt hơn.

- Plugin bảo mật dành cho TMĐT

Đây là một giải pháp với các doanh nghiệp thương mại điện tử quy mô nhỏ, việc sử dụng các plugin bảo mật cho website sẽ giúp tiết kiệm chi phí hơn so với các giải pháp tốn kém khác. Một vài plugin bảo mật e-commerce phổ biến: iThemes Security, Sucuri, Astra.

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 Thương mại điện tử như đã nói ở trên, đồng nghĩa với việc các mặt hàng sản phẩm được bày bán trên các trang thương mại điện tử ngày càng nhiều. Giờ đây từ các thiết bị điện tử có giá trị lớn như laptop, điện thoại thông minh, đồng hồ thông minh, máy tính bảng, ...cho đến các mặt hàng gia nhỏ nhất như móc chìa khóa, nhãn vở, tẩy, bút,... đều được rao bán online. 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. Khi khách hàng đăng tải hình ảnh sản phẩm nào đó muốn mua, hệ thống sẽ nhận biết được đó là sản phẩm gì, thuộc loại gì. Từ đó có thể hiện từ khóa tìm kiếm gợi

ý cho khách hàng quảng cáo những sản phẩm và loại sản phẩm tương tự. Và với điều đó, sản phẩm sẽ đến được tay khách hàng gần hơn, tạo hứng thú mua hàng của khách, thúc đẩy doanh số bán 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. Chúng cho kết quả hơn hẳn so với các kỹ thuật trước kia. Vậy nên việc áp dụng các mô hình học sâu vào mọi lĩnh vực đang ngày càng phổ biến từ y tế, môi trường, nông nghiệp, công nghiệp, tài chính và đương nhiên không ngoại trừ lĩnh vực thương mại, đặc biệt là thương mại điện tử. 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. Trình bày chi tiết về mặt khái niệm, hình thức, đặc trưng của thương mại điện tử và ưu và nhược điểm chức năng và lợi ích của nó đến doanh nghiệp. 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 khái quát ứng dụng mô hình học sâu vào bài toán nhận dạng hình ảnh trong thương mại điện tử.

Các mô hình học sâu, CNN và ứng dụng vào bài toán nhận diện hình ảnh sẽ được trình bày chi tiết ở Chương II dưới đây.

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.

Ví dụ, với một tập các hình ảnh, mỗi hình ảnh mô tả một con mèo hoặc một con chó, thuật toán sẽ “quan sát” toàn bộ dữ liệu và dựa trên hình dạng, màu sắc để hình thành giả thuyết liên quan đến nội dung của ảnh. Kết quả thu được là từ tập dữ liệu ban đầu, các hình ảnh chó/mèo đã được phân loại một cách tự động.

Thực tế, thị giác góp phần tạo nên 80-85% nhận thức của con người về thế giới. Hàng ngày, mỗi người phải thực hiện phân loại trên bất kỳ dữ liệu hình ảnh nào mà chúng ta bắt gặp. Do đó, mô phỏng nhiệm vụ phân loại với sự trợ giúp của mạng nơ-ron là một trong những ứng dụng đầu tiên của thị giác máy tính mà các nhà nghiên cứu nghĩ đến.



Hình 2.1. Minh họa hệ thống phân loại chó và mèo

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. Có 2 định nghĩa về Machine Learning được cung cấp. Theo Arthur Samuel mô tả: “Lĩnh vực nghiên cứu mang lại cho máy tính khả năng học hỏi mà không cần được lập trình rõ ràng.” Đây là một định nghĩa cũ, không chính thức. Tom Mitchell đưa ra một định nghĩa hiện đại và rõ ràng hơn:

“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 .”

Ví dụ: chơi cờ caro.

E = kinh nghiệm chơi nhiều ván cờ caro

T = nhiệm vụ chơi cờ caro.

P = xác suất chương trình sẽ thắng trong trò chơi tiếp theo.

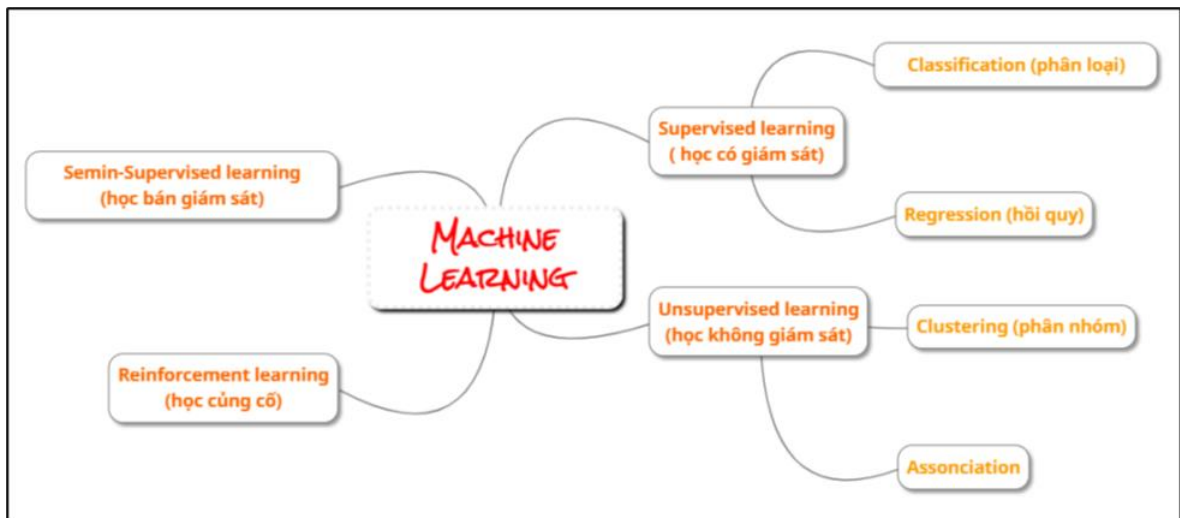
Theo phân nhóm dựa theo phương thức học, Machine learning thường được chia thành 4 loại:

Học có giám sát (Supervised learning): thuật toán dự đoán đầu ra (outcome) của một dữ liệu mới (new input) dựa trên các cặp (input, outcome) đã biết từ trước.

-Học không giám sát (Unsupervised learning): chỉ có dữ liệu vào X mà không biết label Y tương ứng. Chúng ta không biết được đầu ra hay nhãn mà chỉ có dữ liệu đầu vào. Thuật toán unsupervised learning sẽ dựa vào cấu trúc của dữ liệu để thực hiện một công việc nào đó.

-Học bán giám sát (Semi-Supervised Learning): chúng ta có một lượng lớn dữ liệu X nhưng chỉ một phần trong chúng được gán nhãn.

-Học củng cố (Reinforcement Learning): các bài toán giúp cho một hệ thống tự động xác định hành vi dựa trên hoàn cảnh để đạt được lợi ích cao nhất (maximizing the performance).



Hình 2.2. Sơ đồ phân nhóm thuật toán Machine learning

Trong đó bài toán nhận dạng hình ảnh thuộc nhóm thuật toán phân loại và thuộc loại thuật toán học máy là học có giám sát. Một bài toán được gọi là phân loại nếu các nhãn của dữ liệu đầu vào được chia thành một số hữu hạn lớp. Ở đây, tập ảnh gồm các loại đối tượng khác nhau, chia thành các lớp riêng biệt được đánh nhãn.

2.1.2.2 Học sâu

Deep Learning (học sâu) là một loại máy học đặc biệt bắt chước cách tiếp cận học tập mà con người sử dụng để thu thập kiến thức. Neural Network là một phương pháp học sâu. Các model Deep Learning tự nó đã kết hợp 2 phần này lại, các lớp ở phần đầu của model được gọi là Feature Extractor và phần còn lại là để Classify/Regress các features từ phần Feature Extractor để tạo ra kết quả. Do đó phần Feature Extractor này có thể lấy ra được những features từ trong dataset một cách tự động trong quá trình học mà không cần con người định nghĩa các features như trong phương pháp Feature Engineering.

Có nhiều lý do khiến cho các Deep Network hiện đại hiệu quả hơn các phương pháp cổ điển như trên (không kể đến các nguyên nhân chung như lượng dataset lớn hay khả năng tính toán song song, ...), bao gồm:

- Deep Network là end-to-end trainable model: điều này cho phép các Feature Layers tự điều chỉnh những features mà nó cần trích xuất để phù hợp với các

nhiệm vụ tương ứng, trong khi các cách cổ điển chỉ trích xuất các đặc trưng theo ý nghĩ của con người, một cách cố định.

- Khả năng mã hóa: Các bạn nào học Autoencoder cũng đã biết, các Layers đầu sẽ mã hóa dataset thành một tập latent variables và điều này có được là do Feature Layers đã lọc ra những features cần thiết và mã hóa nó (nói ngắn gọn là nén nó) thành một tập dữ liệu nhỏ hơn mà chỉ Decoder Layers của cùng Network mới giải được. Tương tự cho các Feature Extractor khác, nó có nhiệm vụ mã hóa input thành một mẫu phù hợp cho các layers tiếp theo, khi cần thiết nó sẽ thay đổi để đảm bảo những layers kế sẽ nhận input tốt nhất.

Mạng nơ-ron nhân tạo (Neural Network)

Mạng nơ-ron 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.

Khái niệm về mạng lưới thần kinh nhân tạo được lấy cảm hứng từ sinh học của con người và cách các tế bào thần kinh của não người hoạt động cùng nhau để hiểu đầu vào từ các giác quan của con người.

Mạng nơ-ron chỉ là một trong nhiều công cụ và cách tiếp cận được sử dụng trong các thuật toán học máy. Bản thân mạng nơ-ron có thể được sử dụng như một phần trong nhiều thuật toán học máy khác nhau để xử lý đầu vào dữ liệu phức tạp thành một không gian mà máy tính có thể hiểu được.

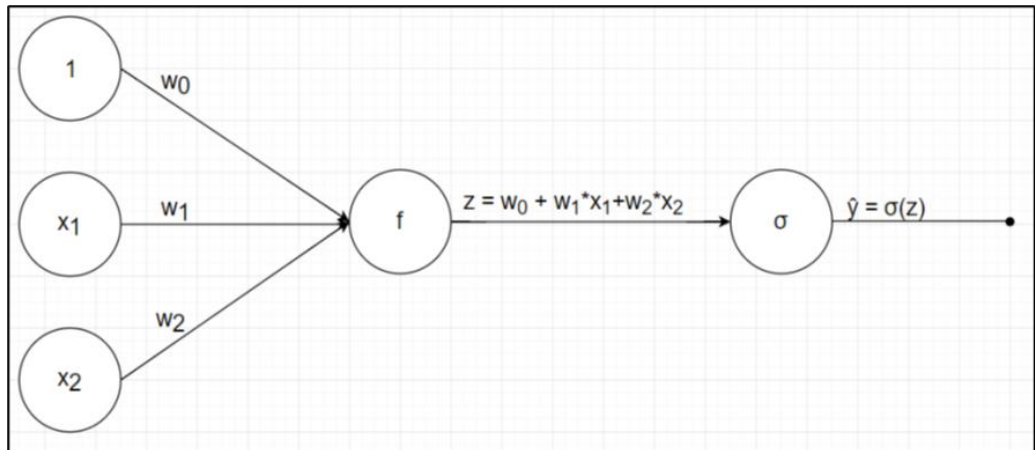
Mạng nơ-ron đang được áp dụng cho nhiều vấn đề trong cuộc sống thực ngày nay, bao gồm nhận dạng giọng nói và hình ảnh, lọc email spam, tài chính và chẩn đoán y tế, ...

Hồi quy Logic (Logistic regression)

Logistic regression là mô hình mạng nơ-ron đơn giản nhất với chỉ một lớp đầu vào và một lớp đầu ra. Cụ thể nó có 2 bước chính:

$$+ \text{Tính tổng linear: } \mathbf{z} = \mathbf{1} * \mathbf{w}_0 + \mathbf{x}_1 * \mathbf{w}_1 + \mathbf{x}_2 * \mathbf{w}_2 \quad (1)$$

$$+ \text{Áp dụng sigmoid function: } \hat{\mathbf{y}} = \sigma(\mathbf{z}) \quad (2)$$



Hình 2.3. Mô hình hồi quy Logic

Hệ số w_0 được gọi là bias. Việc thêm hệ số bias rất quan trọng. Nếu không có hệ số bias thì phương trình đường thẳng chỉ có dạng $y = w_1 * x$, sẽ luôn đi qua gốc tọa độ và nó không tổng quát hóa phương trình đường thẳng nên sẽ không tìm được phương trình mong muốn.

Hàm sigmoid ở đây được gọi là kích hoạt (activation function)

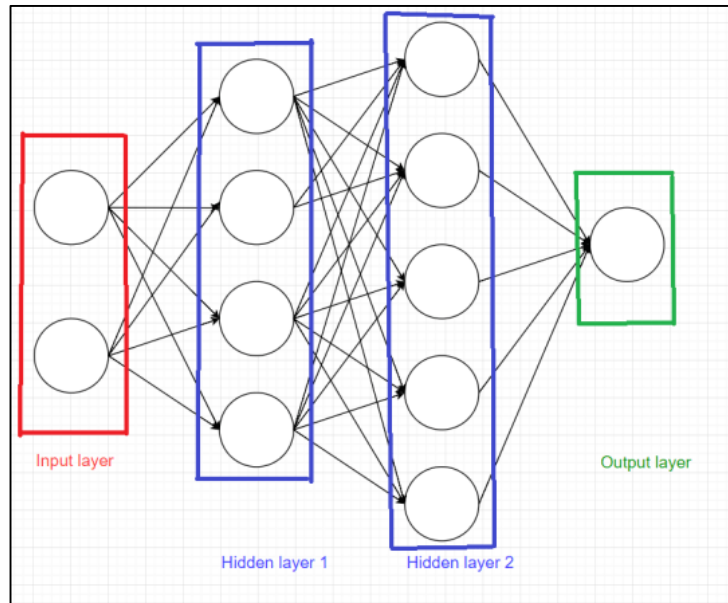
Mô hình tổng quát

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 2.4 Mô hình mạng neural network

- **Cách hoạt động**

Các thuật toán học máy sử dụng mạng nơ-ron thường không cần được lập trình với các quy tắc cụ thể xác định những gì mong đợi từ đầu vào. Thay vào đó, thuật toán học mạng nơ-ron học từ việc xử lý nhiều ví dụ được gắn nhãn (tức là dữ liệu có "câu trả lời") được cung cấp trong quá trình đào tạo và sử dụng khóa trả lời này để tìm hiểu đặc điểm nào của đầu vào là cần thiết để tạo ra đầu ra chính xác.

Khi đã xử lý đủ số lượng ví dụ, mạng nơ-ron có thể bắt đầu xử lý các đầu vào mới, chưa nhìn thấy và trả về thành công kết quả chính xác. Chương trình càng thấy nhiều ví dụ và nhiều loại đầu vào thì kết quả thường càng chính xác hơn vì chương trình học hỏi bằng kinh nghiệm.

2.1.2.3 Mạng nơ-ron tích chập (CNN)

Convolutional Neural Network (CNN – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến giúp xây dựng được những hệ thống thông minh với độ chính xác cao hiện nay như hệ thống lớn xử lý ảnh Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay máy giao hàng tự động... CNN được sử dụng nhiều trong các bài toán nhận dạng các đối tượng trong ảnh.

Về cơ bản CNN 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).

- **Lấy chập (Convolution)**

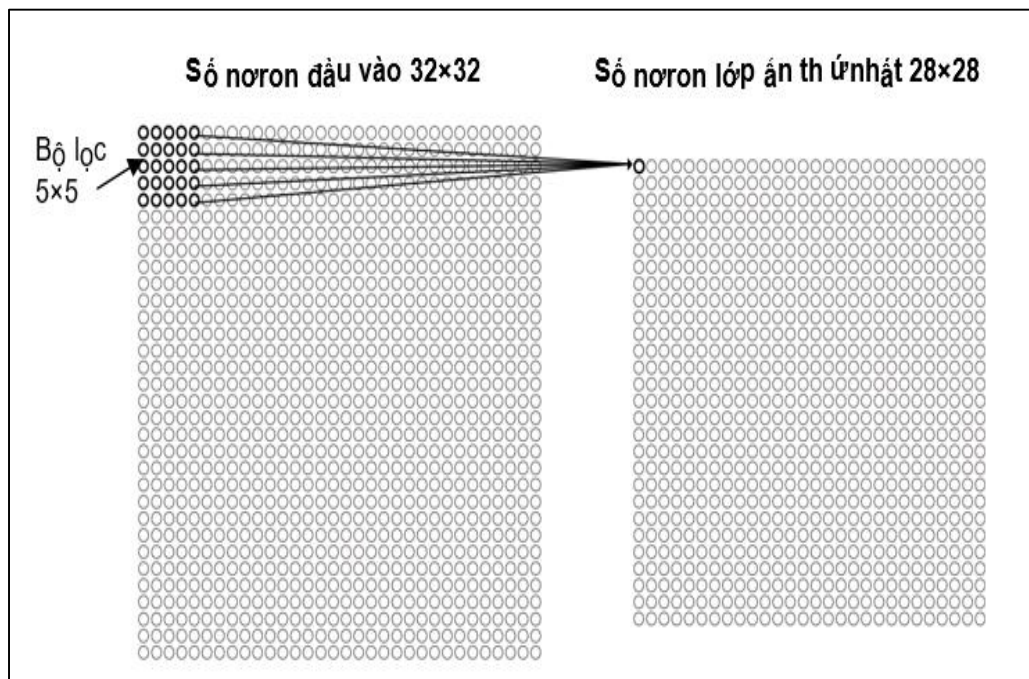
Là thao tác đầu tiên quan trọng nhất trong cấu trúc của mạng học sâu CNN. Phép lấy chập dựa trên lý thuyết xử lý tín hiệu số, thực hiện các xử lý về mặt toán học tính lấy chập để giúp trích xuất được những thông tin quan trọng từ dữ liệu. Đầu vào của phép lấy chập là một mảng các giá trị của dữ liệu. Chẳng hạn, trong phân loại ảnh, đầu vào là một ảnh được biểu diễn bằng mảng $32 \times 32 \times 3$ các giá trị pixel (mỗi phần tử của mảng có giá trị trong khoảng từ 0 đến 255 biểu diễn cường độ sáng của pixel tại một điểm ảnh).

Để thực hiện lấy chập, một bộ lọc (Filter) còn gọi là Kernel được di chuyển qua các vị trí trên toàn bộ ma trận ảnh. Bộ lọc này thực chất là một cửa sổ có kích thước $n \times n$ (kí hiệu $F = n$) chứa các số (các số này chính là trọng số hay tham số). Kích thước của bộ lọc thường là nhỏ (chẳng hạn 3×3 hoặc lớn nhất là 5×5). Bộ lọc sẽ di chuyển từ trái qua phải, từ trên xuống dưới với bước dịch chuyển $S=1$ cho cả hai chiều, vị trí đầu tiên của bộ lọc là góc trên bên trái. Thao tác lấy chập được thực hiện tại các vị trí mà bộ lọc đi qua. Ý nghĩa của thao tác lấy chập là xác định khả năng xuất hiện các mẫu tại các vị trí nhất định trong ảnh. Mỗi mẫu được biểu diễn bằng trọng số của cửa sổ tương ứng với một bộ lọc. Mỗi vị trí của bộ lọc sẽ tính được một giá trị theo công thức:

$$y = \sum w_i x_i + b_i \quad (3)$$

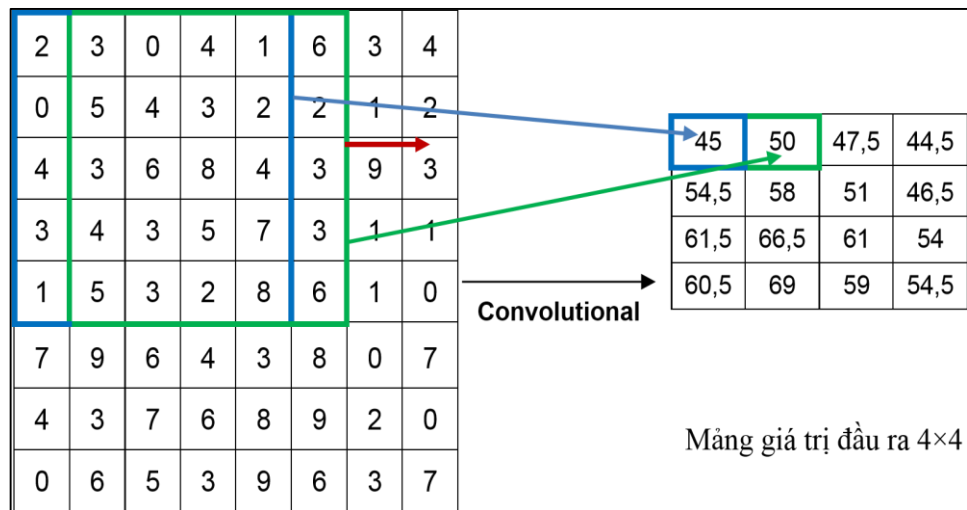
Trong công thức (3), x_i bao gồm các điểm ảnh phổ nằm trong phạm vi cửa sổ đang quét, giả sử kích thước bộ lọc 5×5 sẽ có $i=25$ điểm ảnh phổ được quét. Số giá trị phải quét nằm tại cùng vị trí cho tất cả các thành phần, do vậy sẽ có $K \times 5 \times 5$ giá trị x_i và w_i tương ứng. Ngoài ra, còn thêm một hệ số độ lệch b trong công thức này. Do vậy, số tham số cần thiết cho thao tác lấy chập bao gồm w_i và b là $K \times 5 \times 5 + 1$ tham số. Giả thiết dùng M bộ lọc, số lượng tham số sẽ là $M \times (K \times 5 \times 5 + 1)$ tham số.

Sau khi trượt bộ lọc qua tất cả các vị trí và lấy chập sẽ được một mảng với mỗi giá trị là các số được tính bằng cách lấy chập theo công thức trên. Đối với ví dụ mảng đầu vào của ảnh ở trên ta được mảng $28 \times 28 \times 1$ các giá trị. Mảng này được gọi là ánh xạ kích hoạt (Activation Map) hay ánh xạ đặc trưng (Feature Map). Lý do có mảng 28×28 vì có 784 vị trí khác nhau để bộ lọc 5×5 có thể khớp trên ảnh 32×32 , 784 giá trị này được ánh xạ thành mảng 28×28 . Hình 7 mô tả bước lấy chập cho ví dụ ảnh đầu vào kích thước 32×32 , đầu ra là ánh xạ đặc trưng kích thước 28×28 .



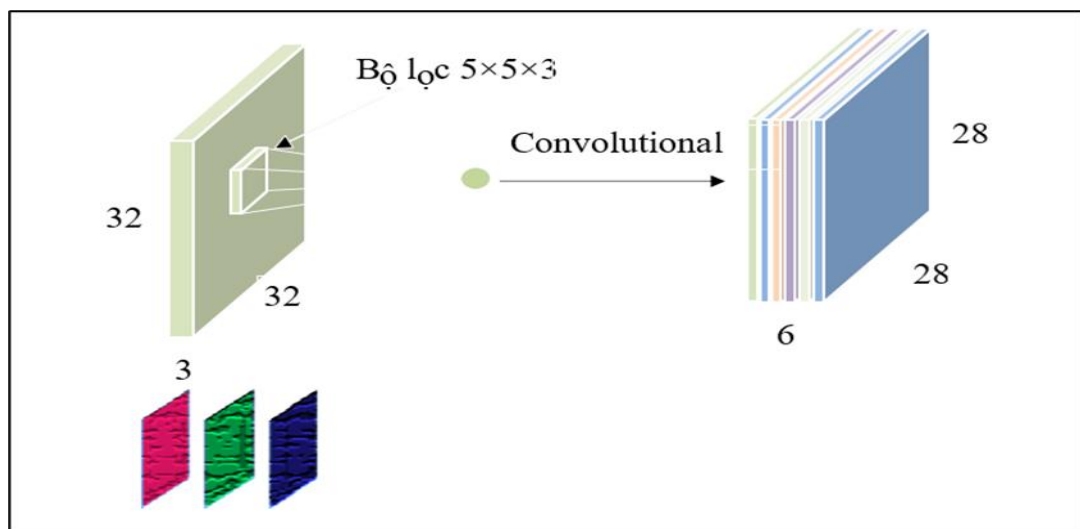
Hình 2.5. Mô tả lấy chập dùng bộ lọc kích thước 5×5

Hình dưới đây mô tả lấy chập trên một phần ảnh đầu vào với mảng giá trị có kích thước 8×8 . Bộ lọc dịch chuyển 5×5 sẽ di chuyển qua toàn bộ phần ảnh và thao tác lấy chập được áp dụng cho 25 nút trên cửa sổ bộ lọc. Giả thiết ma trận trọng số của bộ lọc có các giá trị đều là 0,5. Kết quả thu được ở đầu ra là một mảng ánh xạ đặc trưng có kích thước 4×4 . Với mỗi ô vuông màu xanh tượng trưng cho cửa sổ bộ lọc sẽ được tính toán (lấy chập) cho ra một giá trị tương ứng ở đầu ra. Bước trượt sẽ kiểm soát bộ lọc lấy chập khắp ảnh vào. Mặc định, mỗi lần bộ lọc sẽ dịch chuyển đi một đơn vị. Số lượng đơn vị mà bộ lọc dịch đi được gọi là bước trượt (stride).



Hình 2.6. Mô tả chi tiết lấy chập dùng bộ lọc kích thước 5x5

Chú ý rằng, các tham số của bộ lọc (trọng số) được giữ nguyên khi bộ lọc di chuyển qua ảnh đầu vào. Điều này cho phép bộ lọc cần được huấn luyện nhận ra một số đặc trưng trong dữ liệu vào. Đối với nhận dạng ảnh, có thể học để nhận biết các hình dạng như đoạn thẳng, cạnh và các hình dạng phân biệt khác. Vì thế bước lấy chập còn được gọi là ánh xạ đặc trưng. Tuy nhiên, để phân loại tốt, ở mỗi tầng lấy chập thường cần nhiều bộ lọc. Ví dụ, Hình 2.7 sử dụng 3 bộ lọc có kích thước 5×5 với đầu vào là ảnh màu kích thước 32×32, đầu ra của lấy chập là ánh xạ đặc trưng xếp chồng có kích thước 28×28×6.



Hình 2.7. Mô tả bước lấy chập của mạng nơron dùng 3 bộ lọc kích thước 5×5

Ưu điểm của lấy chập:

Giảm số lượng tham số: Ở NN truyền thống, các nơon ở lớp trước sẽ kết nối tới tất cả các nơon ở lớp sau (fully connected) gây nên tình trạng quá nhiều tham số cần học. Đây là nguyên nhân chính gây nên tình trạng quá khớp (overfitting) cũng như làm tăng thời gian huấn luyện. Việc sử dụng lấy chập trong đó cho phép chia sẻ trọng số liên kết (shared weights), cũng như thay vì sử dụng kết nối đầy đủ (fully connected) sẽ sử dụng trường tiếp nhận cục bộ (local receptive fields) giúp giảm tham số.

Các tham số trong quá trình sử dụng lấy chập hay giá trị của các filter - kernel sẽ được học trong quá trình huấn luyện.

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

Về cơ bản, lấy chập là một phép biến đổi tuyến tính. Nếu tất cả các nơon được tổng hợp bởi các phép biến đổi tuyến tính thì một mạng nơon 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.

Một số hàm kích hoạt phi tuyến thường dùng như ReLU (Rectified Linear Unit), ELU (Exponential Linear Unit). ReLU có hàm kích hoạt dạng $f(x) = \max(0, x)$ cho các giá trị vào. Về cơ bản, hàm này sẽ thay đổi tất cả các giá trị kích hoạt âm thành 0 và tăng tính phi tuyến của mô hình và toàn mạng mà không ảnh hưởng tới lớp lấy chập. Hàm ELU có dạng như dưới với $\alpha > 0$:

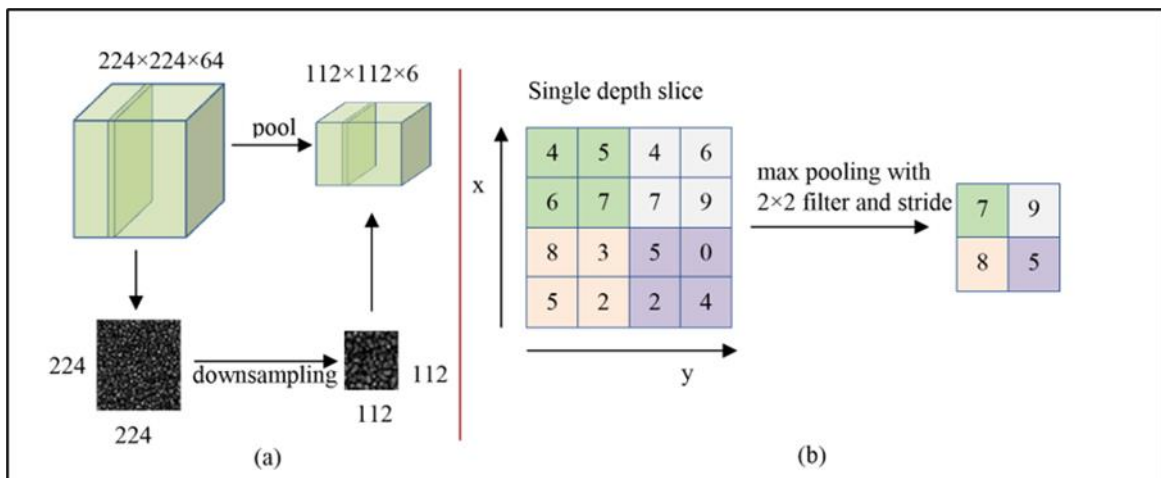
$$f(x) = \begin{cases} x & \text{nếu } x \geq 0 \\ \alpha (\exp(x) - 1) & \text{nếu } x < 0 \end{cases} \quad (4)$$

Hàm ReLU thực hiện tính toán đơn giản hơn ELU. Tuy nhiên, các nghiên cứu gần đây cho thấy việc sử dụng hàm ELU cho hiệu suất tốt hơn.

- **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. Điều này cũng tương tự như là bộ lọc phát hiện vị trí đối tượng trong bài toán phát hiện đối tượng trong ảnh. Nhìn chung, bộ lọc di chuyển thường có kích thước $F = 2$ và bước trượt $S = 2$ được dùng phổ biến. Có ít các thiết lập sử dụng bộ lọc di chuyển có kích thước $F = 3$ và $S = 2$. Và hiếm có thiết lập kích thước lớn hơn 3. Do kích thước bộ lọc di chuyển quá lớn có thể sẽ dẫn đến mất mát một số thông tin hữu ích, điều này làm cho hiệu suất nhận dạng kém đi.

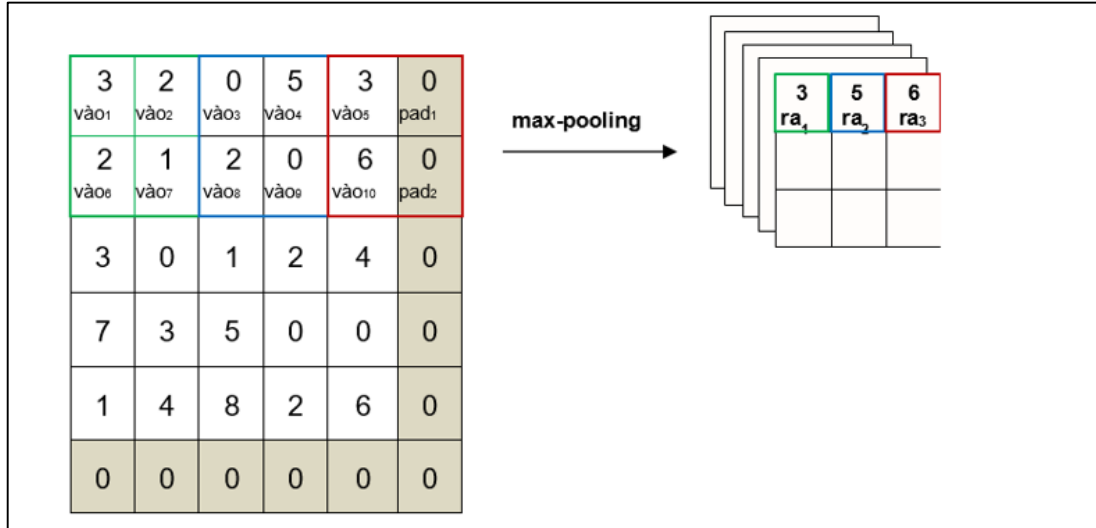


Hình 2.8. Ví dụ sử dụng max-pooling

Hình 2.8 (a) là cách thức pooling xử lý đối với một đầu vào là kết quả của nhiều bộ lọc ($k = 64$), kích thước của đầu vào là $224 \times 224 \times 64$ được thực hiện với $F = 2$, $S = 2$, đầu ra sẽ có kích thước $112 \times 112 \times 64$. Hình 2.8(b) mô tả chi tiết cách thức hoạt động của max-pooling trong đó $F = 2$, $S = 2$ và kết quả đầu ra ma trận tương ứng.

Trong quá trình lấy gộp, để bảo toàn thông tin cho ảnh đầu vào thường sử dụng padding với giá trị bằng 0 để không ảnh hưởng đến quá trình lấy gộp trên ảnh đầu

vào. Hình 2.9 mô tả cách thực hiện max-pooling với padding và bộ lọc $F = 2$ cho toàn bộ ảnh.

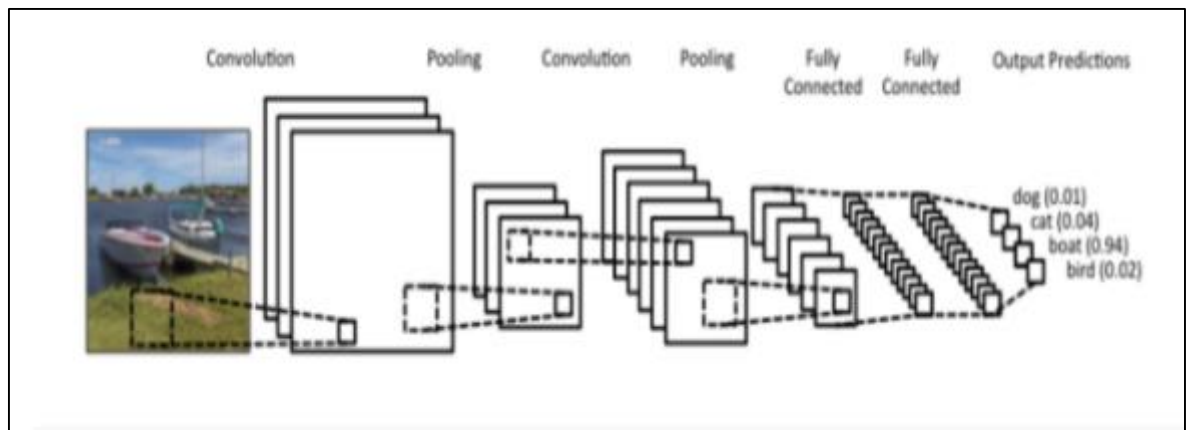


Hình 2.9. Mô tả cách thực hiện max-pooling với padding

- **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.

Như vậy, CNN là giải thuật có kiến trúc bao gồm nhiều tầng có chức năng khác nhau trong đó tầng chính hoạt động thông qua cơ chế lấy chập. Trong suốt quá trình huấn luyện, CNN sẽ tự động học được các thông số cho các bộ lọc tương ứng là các đặc trưng theo từng cấp độ khác nhau. Ví dụ trong bài toán phân lớp ảnh, CNN sẽ cố gắng tìm ra các thông số tối ưu cho các bộ lọc tương ứng theo thứ tự điểm ảnh \rightarrow các cạnh \rightarrow khuôn hình \rightarrow bộ mặt \rightarrow các đặc trưng mức cao (pixel \rightarrow edges \rightarrow shapes \rightarrow facial \rightarrow high-level features). Đây chính là lý do mà CNN có được kết quả vượt trội so với các giải thuật trước đây.



Hình 2.10. Mô hình mạng neural tích chập

2.2 Một số nghiên cứu có liên quan

2.2.1 Quá trình phát triển của các kiến trúc CNN

Tiến trình phát triển của các kiến trúc CNN có thể được khái quát qua những dấu mốc quan trọng sau đây:

- 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. Tuy nhiên vào thời điểm đó do chưa có sự phát triển của dữ liệu và khả năng tính toán nên mạng CNN vẫn chưa có cơ hội bùng nổ. Các mô hình machine learning truyền thống như SVM, KNN vẫn được sử dụng phổ biến.

- 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. 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. Các mô hình được huấn luyện trên ImageNet có thể chuyển giao tới rất nhiều những domain dữ liệu khác nhau. Kể từ thời điểm 2010, ImageNet trở thành tiêu chuẩn đo đạc 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. Dấu mốc đó đã khởi đầu cho xu hướng ứng dụng CNN trong computer vision thay thế cho những thuật toán học máy truyền thống trước kia.

Liên tiếp vào những năm sau đó, ngày càng xuất hiện nhiều các kiến trúc CNN mới. Chúng được hình thành, phát triển và cải tiến về độ sâu, cách thiết kế block, cách kết nối giữa các block. Lần lượt từ VGG Net, GoogleNet, ResNet, DenseNet ... mà chúng ta sẽ tìm hiểu qua bài viết này được ra đời dựa trên sự kế thừa những ý tưởng cũ và phát triển những ý tưởng mới mẻ. Quá trình phát triển của các kiến trúc mạng song hành cùng với sự phát triển phần cứng máy tính như các GPU có tốc độ nhanh hơn. Kỹ thuật huấn luyện phân tán và song song trên nhiều GPU cho phép một model huấn luyện chỉ trong vòng một vài tiếng so với việc huấn luyện kéo dài qua nhiều ngày và tốn kém như trước đây. Các framework hỗ trợ deep learning cũng xuất hiện nhiều hơn, được cải tiến và trở thành công cụ đáp ứng mọi nhu cầu cần thiết cho quá trình huấn luyện deep learning. Phổ biến nhất có thể kể tới ba frameworks pytorch (facebook), tensorflow (google), mxnet (intel) được phát triển và hậu thuẫn từ những công ty công nghệ hàng đầu thế giới. Kể từ sau ImageNet, các bộ dữ liệu ảnh đã khẳng định vai trò thúc đẩy sự phát triển của ngành AI. Các thuật toán được so sánh với nhau dựa trên kết quả dẫn đầu (leader board) từ những bộ dữ liệu chuẩn hoá. Nhờ sự mở rộng của những nền tảng huấn luyện free như google colab, kaggle mà mọi người đều có thể tiếp cận được với AI. Chiến lược phát triển toàn cầu về AI ở các tập đoàn, quốc gia trên thế giới dẫn tới sự hình thành những viện nghiên cứu về AI quy tụ được nhiều nhà khoa học xuất sắc và có những nghiên cứu đột phá.

2.2.2. Các mạng CNN tiêu biểu

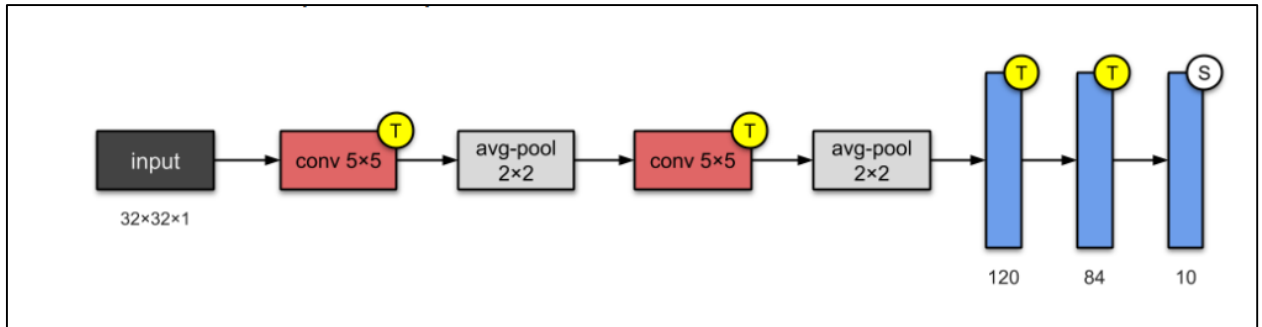
- **LeNet-5 (1998)**

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.

Trong kiến trúc mạng nơ ron đầu tiên, để giảm chiều dữ liệu, Yan Lecun sử dụng Sub-Sampling Layer là một Average-Pooling Layer (các layer nhằm mục đích giảm

chiều dữ liệu mà không thay đổi đặc trưng chúng ta còn gọi là Sub-Sampling Layer). Kiến trúc này khó hội tụ nên ngày nay chúng được thay thế bằng Max-Pooling.

Đầu vào của mạng LeNet có kích thước nhỏ (chỉ 32×32) và ít layers nên số lượng tham số của nó chỉ khoảng 60 nghìn.



Hình 2.11. Kiến trúc LeNet

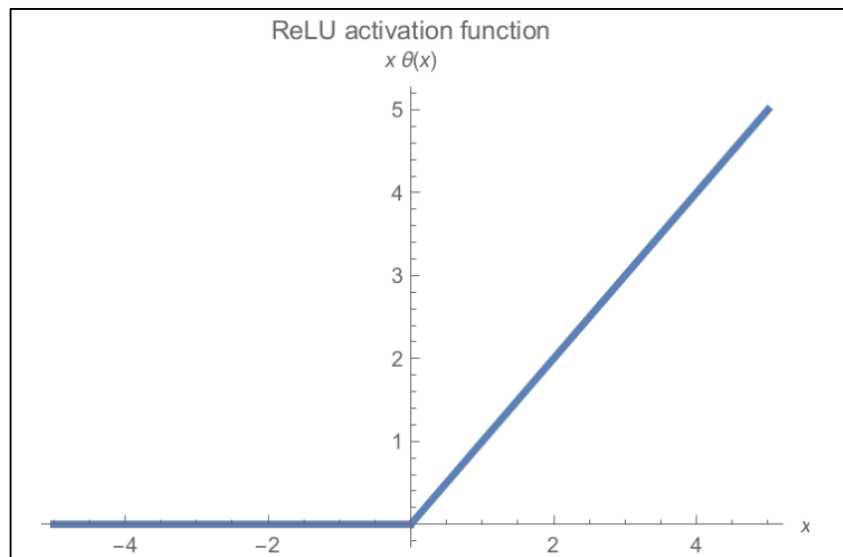
- **AlexNet (2012)**

AlexNet là mạng CNN được giới thiệu vào năm 2012 bởi Alex Krizhevsky và giành chiến thắng trong cuộc thi ImageNet với cách biệt khá lớn so với vị trí thứ hai. Lần đầu tiên Alex net đã phá vỡ định kiến trước đó cho rằng các đặc trưng được học từ mô hình sẽ không tốt bằng các đặc trưng được tạo thủ công (thông qua các thuật toán SUFT, HOG, SHIFT). 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.

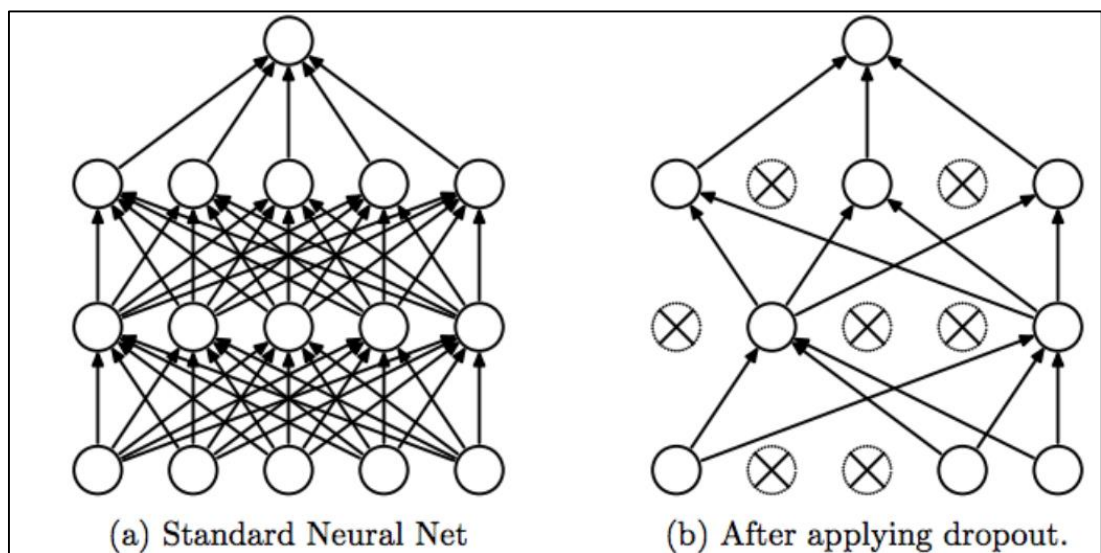
Ngoài ra mạng còn cải tiến trong quá trình optimizer như:

- Lần đầu tiên sử dụng activation là ReLU (Rectified Linear Unit) thay cho Sigmoid. ReLU là hàm có tốc độ tính toán nhanh nhờ đạo hàm chỉ có 2 giá trị $\{0, 1\}$ và không có lũy thừa cơ số như hàm sigmoid nhưng vẫn tạo ra được tính phi tuyến (non-linear).



Hình 2.12. Hàm ReLu

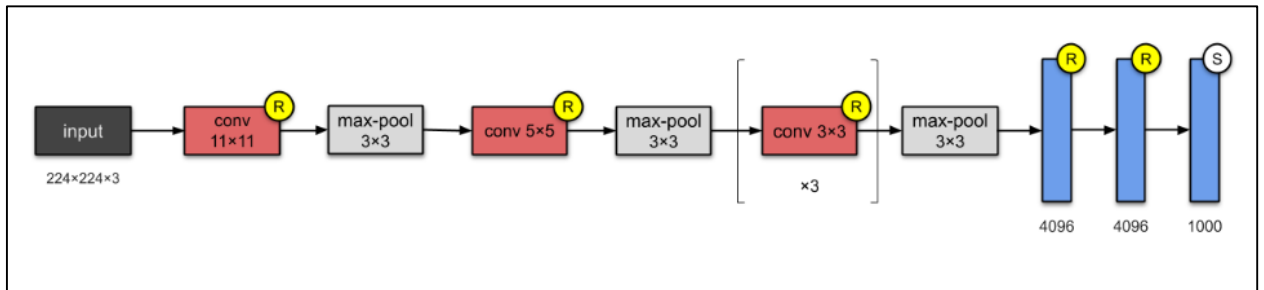
▪ Sử dụng dropout layer giúp giảm số lượng liên kết neural và kiểm soát overfitting. Phương pháp dropout có số lượng các liên kết mạng bị giảm so với trước đó làm mô hình ít phức tạp hơn. Đồng thời đây cũng là một dạng ensemble model giúp giảm thiểu được overfitting.



Hình 2.13. 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 2.14. Kiến trúc mạng AlexNet

- **VGG-16 (2014)**

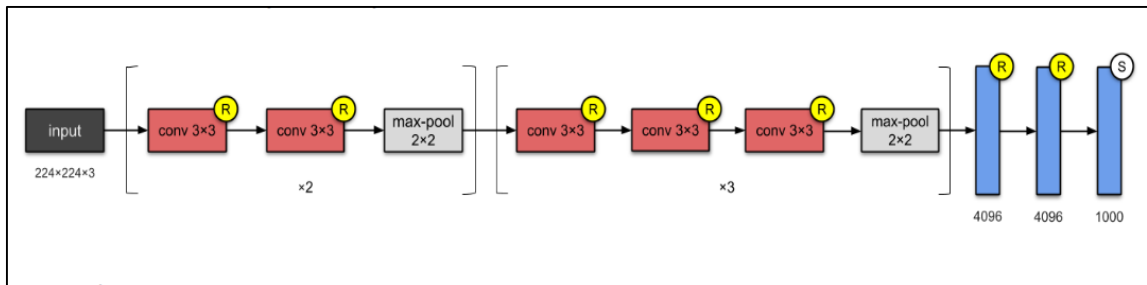
Với VGG-16, quan điểm về một mạng nơ ron sâu hơn sẽ giúp ích cho cải thiện độ chính xác của mô hình tốt hơn. Về kiến trúc thì VGG-16 vẫn giữ các đặc điểm của AlexNet nhưng có những cải tiến:

- Kiến trúc VGG-16 sâu hơn, bao gồm 13 layers tích chập 2 chiều (thay vì 5 so với AlexNet) và 3 layers fully connected.
- Lần đầu tiên trong VGG-16 chúng ta xuất hiện khái niệm về khối tích chập (block). Đây là những kiến trúc gồm một tập hợp các layers CNN được lặp lại giống nhau. Kiến trúc khối đã khởi nguồn cho một dạng kiến trúc hình mẫu rất thường gặp ở các mạng CNN kể từ đó.
- VGG-16 cũng kế thừa lại hàm activation ReLU ở AlexNet.
- VGG-16 cũng là kiến trúc đầu tiên thay đổi thứ tự của các block khi xếp nhiều layers CNN + max pooling thay vì xen kẽ chỉ một layer CNN + max pooling. Các layers CNN sâu hơn có thể trích lọc đặc trưng tốt hơn so với chỉ 1 layers CNN.
- VGG-16 chỉ sử dụng các bộ lọc kích thước nhỏ 3x3 thay vì nhiều kích thước bộ lọc như AlexNet. Kích thước bộ lọc nhỏ sẽ giúp giảm số lượng tham số cho mô hình và mang lại hiệu quả tính toán hơn. VD: Nếu sử dụng 2 bộ lọc kích thước 3 x 3 trên một features map (là output của một layer CNN) có độ sâu là 3 thì ta sẽ cần $n_filters \times kernel_size \times kernel_size \times n_channels = 2 \times 3 \times 3 \times$

3 = 54 tham số. Nhưng nếu sử dụng 1 bộ lọc kích thước 5×5 sẽ cần $5 \times 5 \times 3 = 75$ tham số. 2 bộ lọc 3×3 vẫn mang lại hiệu quả hơn so với 1 bộ lọc 5×5 .

Mạng VGG-16 sâu hơn so với AlexNet và số lượng tham số của nó lên tới 138 triệu tham số. Đây là một trong những mạng mà có số lượng tham số lớn nhất. Kết quả của nó hiện đang xếp thứ 2 trên bộ dữ liệu ImageNet validation ở thời điểm public. Ngoài ra còn một phiên bản nữa của VGG-16 là VGG-19 tăng cường thêm 3 layers về độ sâu.

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 2.15. 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.

Những Pretrained Model như vậy thường được train trên một hoặc một vài bộ datasets nhất định, tương thích và cho accuracy cao với một nhiệm vụ hoặc nhiều nhiệm vụ (multi-task deep learning) nào đó mà nó được train. Chúng ta gọi các nhiệm vụ mà pretrained model đó được train để thực hiện là source tasks.

Nhiệm vụ của chúng ta là tạo ra một model mới để thực hiện một hoặc nhiều tasks nào đó. Những tasks cần được thực hiện của model này có thể trùng hoặc không

trùng với tasks mà pretrained model được train (thường thì sẽ không trùng), chúng ta gọi tasks này là target tasks.

Cụ thể, Transfer Learning trong Deep Learning là một kỹ thuật mà trong đó: Một pretrained model đã được train trên source tasks cụ thể nào đó, khi đó một phần hay toàn bộ pretrained model có thể được tái sử dụng phụ thuộc vào nhiệm vụ của mỗi layer trong model đó.

Một model mới sử dụng một phần hay toàn bộ pretrained model để học một target tasks và tùy vào nhiệm vụ của mỗi layer mà model mới có thể thêm các layer khác dựa trên pretrained model sẵn có.

Việc sử dụng pretrained model là một bước tiến lớn để những người đi sau tiếp bước những thành quả của các bậc tiền bối, tận dụng những pretrained model sẵn có để tạo ra những model mới phục vụ cho các target tasks cụ thể hơn, mang tính ứng dụng thực tiễn hơn.

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. Ví dụ, bỏ lớp đầu của pretrained model đi, lớp kết nối đầy đủ cuối cùng cho kết quả ra những đặc trưng, sau khi lấy ra những đặc trưng của ảnh bằng việc sử dụng ConvNet của pre-trained model, ta cho những đặc trưng ấy vào các mô hình học máy như Support vector machine, Softmax... để phân loại hoặc K-means, Spectral clustering... để phân cụm.

- Fine tuning: Thường bỏ đi lớp đầu ra cuối cùng (output layer) hoặc một phần các lớp trước đó 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. Chúng có các khối dư sẽ được trình bày cụ thể dưới đây. Các khối như vậy được xếp chồng lên nhau từ đầu đến cuối cho phép chúng ta điều chỉnh hình dạng đầu ra nhỏ hơn dần.

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 dốc 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:

$$\hat{x}_i = \frac{x_i - \mu}{\sigma} \quad (5)$$

Với các giá trị μ và σ được tính như sau:

$$\mu = \frac{1}{m} \sum_{i=1}^m \mathbf{x}_i \quad (6)$$

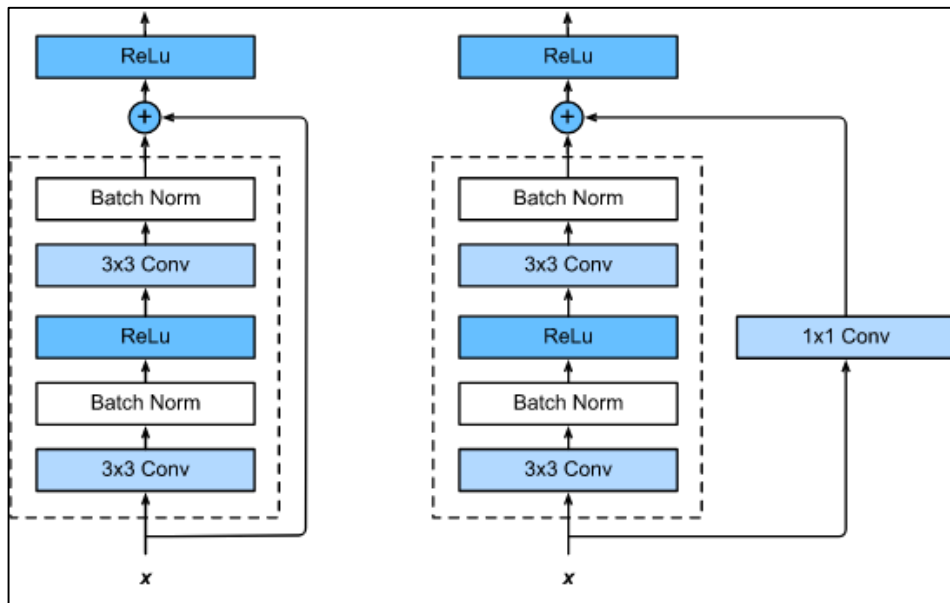
$$\sigma^2 = \sum_{i=1}^m (\mathbf{x}_i - \mu)^2 \quad (7)$$

2.4.3 Kết nối tắt (Skip Connection)

Những kiến trúc trước đây thường cải tiến độ chính xác nhờ gia tăng chiều sâu của mạng CNN. Nhưng thực nghiệm cho thấy đến một ngưỡng độ sâu nào đó thì độ chính xác của mô hình sẽ bão hòa và thậm chí phản tác dụng và làm cho mô hình kém chính xác hơn. Khi đi qua quá nhiều tầng độ sâu có thể làm thông tin gốc bị mất đi

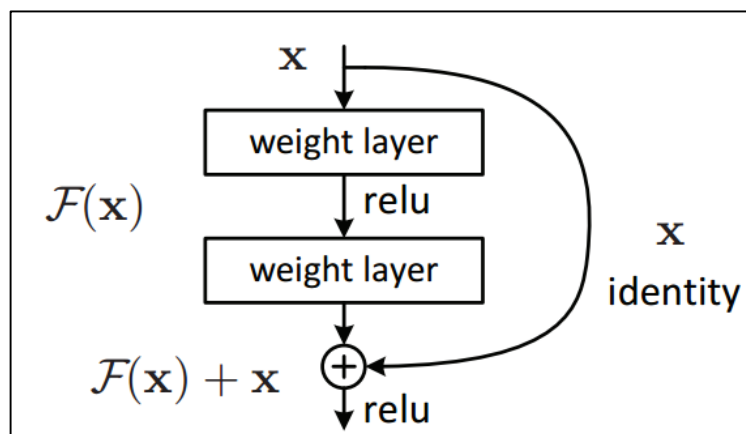
thì các nhà nghiên cứu của Microsoft đã giải quyết vấn đề này trên ResNet bằng cách sử dụng kết nối tắt.

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



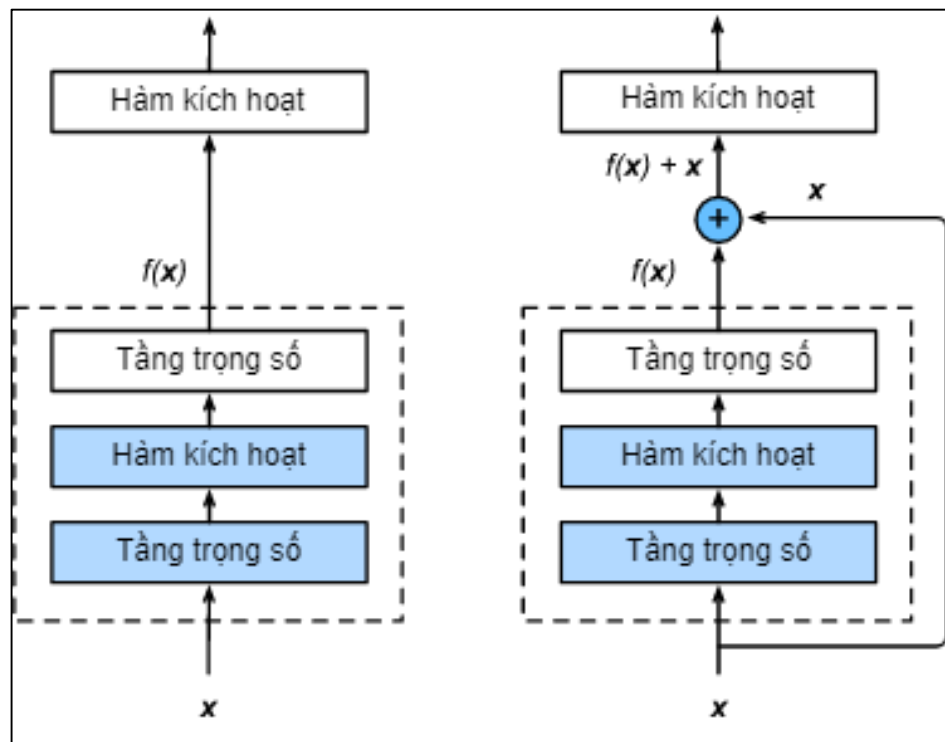
Hình 2.16. Khối ResNet thông thường và khối ResNet với tầng tích chập 1x1

Khối xác định (Identity block) thì không áp dụng tích chập 1 x 1 mà cộng trực tiếp giá trị của nhánh đó vào nhánh còn lại.



Hình 2.17. Khối xác định (Identity block)

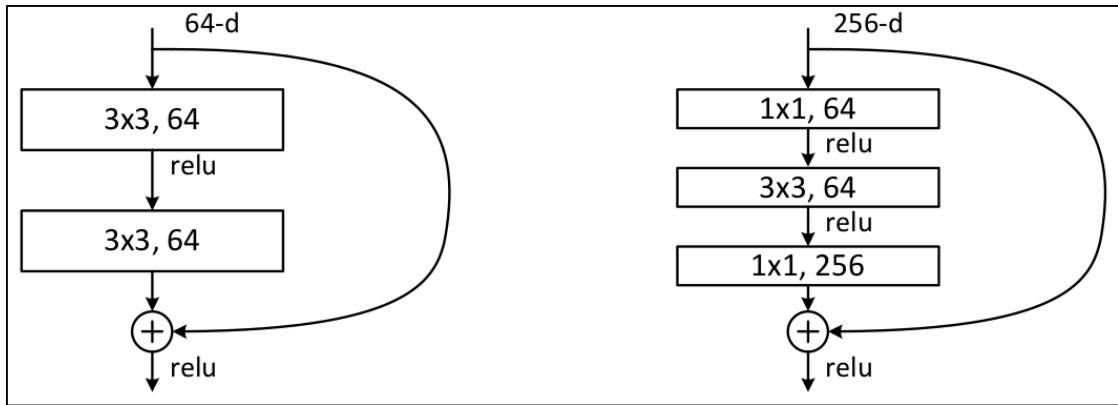
Giả sử chúng ta có x là đầu vào của khối xác định. Chúng ta cần ánh xạ đầu vào x thành hàm $f(x)$. Để tìm ra ánh xạ chuẩn xác tương đương với hàm $f(x)$ là một việc khá khó. Nhưng nếu cộng thêm ở đầu ra thành $x+f(x)$ thì chúng ta sẽ quy về tham số hóa độ lệch, tức cần tham số hóa phần dư $f(x)$. Tìm ánh xạ theo phần dư sẽ dễ hơn nhiều vì chỉ cần tìm giá trị $f(x)$ sao cho nó gần bằng 0 là có thể thu được một ánh xạ chuẩn xác. Tại một khối xác định, chúng ta sẽ áp dụng một layer activation ReLU sau mỗi xen kẽ giữa những tầng trọng số.



Hình 2.18. Sự khác biệt giữa một khối thông thường (trái) và một khối xác định(phải)

2.4.4. Mô hình ResNet50 áp dụng vào bài toán

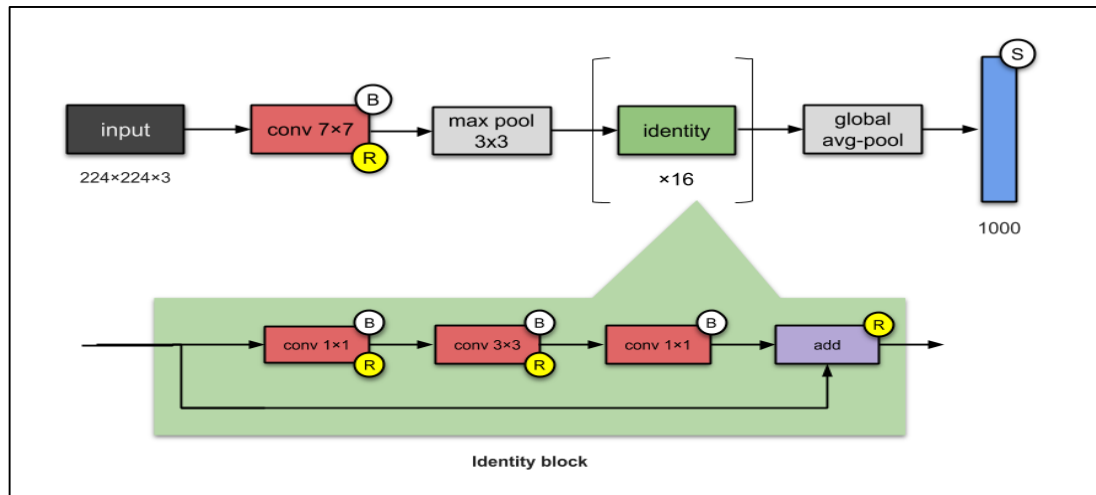
Kiến trúc Resnet50 cũng giống như kiến trúc ResNet ở trên, có một điểm khác biệt chính. So với Resnet34, các kết nối tắt đã qua hai lớp nhưng bây giờ chúng qua ba lớp và cũng có các lớp tích chập 1×1 được thêm vào. Điều này có độ chính xác cao hơn nhiều so với mô hình ResNet 34 lớp. ResNet 50 lớp đạt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



Hình 2.19. Kết nối tắt qua 3 lớp của RestNet50 so với 2 lớp của RestNet34

Kiến trúc cơ sở của RestNet50 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 2.20. Kiến trúc tóm tắt của mạng ResNet50.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Hình 2.21. Kiến trúc chi tiết của RestNet50

Vì vậy, như chúng ta có thể thấy trong hình 2.21, kiến trúc ResNet 50 chứa thành phần như sau:

- Một convoultion với kích thước kernel là $7 * 7$ và 64 kernel khác nhau, tất cả đều có kích cỡ stride là 2 cho chúng ta **1 lớp**.
- Tiếp theo, chúng ta thấy max pooling cũng với kích cỡ stride là 2.
- Trong convoultion tiếp theo, có một kernel $1 * 1,64$ theo sau là kernel $3 * 3,64$ và cuối cùng là một kernel $1 * 1,256$, ba lớp này được lặp lại tổng cộng 3 lần, do đó chúng ta có **9 lớp** tiếp theo.
- Tiếp theo, chúng ta thấy kernel $1 * 1,128$ sau đó là kernel $3 * 3,128$ và cuối cùng là kernel $1 * 1,512$, nó được lặp lại 4 lần để cho ra **12 lớp**.
- Sau đó, có một kernel $1 * 1,256$ và hai hạt nhân nữa với $3 * 3,256$ và $1 * 1,1024$ và điều này được lặp lại 6 lần cho chúng ta tổng cộng **18 lớp**.
- Và sau đó lại là một kernel $1 * 1,512$ với hai kernel nữa là $3 * 3,512$ và $1 * 1,2048$ và nó được lặp lại 3 lần cho chúng ta tổng cộng **9 lớp**.
- Cuối cùng, ResNet50 có average pool và kết thúc nó bằng một lớp kết nối đầy đủ (fully connected layer) chứa 1000 nút và ở cuối có một hàm softmax để có **1 lớp**..

Không đếm hàm kích hoạt và các lớp gộp max/ average pooling nên tổng cộng ResNet50 cho chúng ta một mạng lưới sâu $1 + 9 + 12 + 18 + 9 + 1 = 50$ lớp.

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

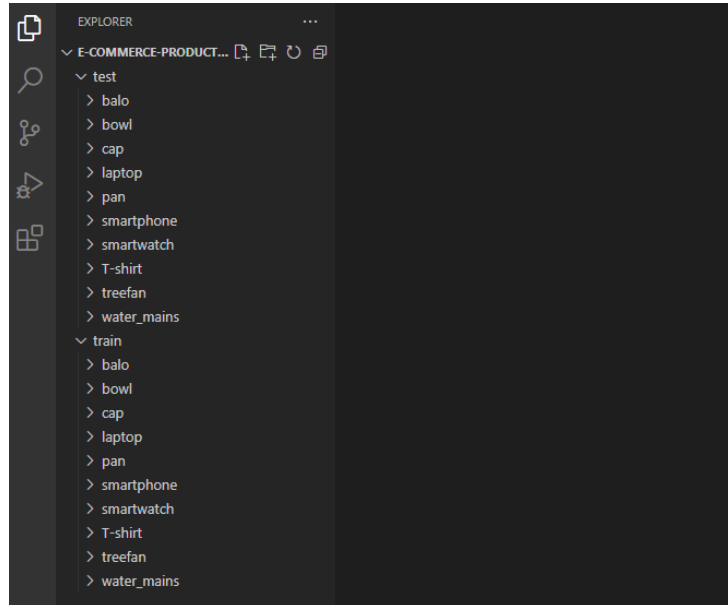
Tập ảnh thử nghiệm là tập ảnh các sản phẩm thương mại điện tử phổ biến gồm các loại mặt hàng khác nhau. Tập dữ liệu gồm 10 sản phẩm, thuộc 3 loại sản phẩm: đồ điện tử, đồ gia dụng, đồ thời trang.

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. Chi tiết tập dữ liệu như sau:

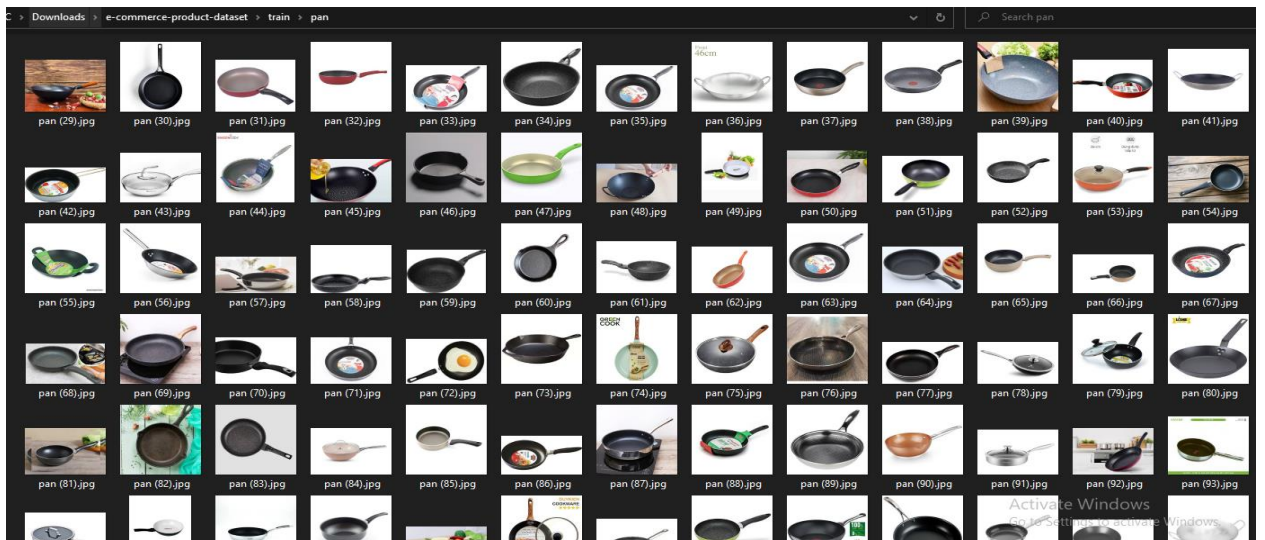
STT	Tên thư mục	Tên sản phẩm	Số lượng ảnh	Loại sản phẩm
1	laptop	Laptop	107	Đồ điện tử
2	smartphone	Điện thoại thông minh	104	
3	smartwatch	Đồng hồ thông minh	109	Đồ gia dụng
4	treefan	Quạt điện cây	103	
5	pan	Chảo	140	
6	water_mans	Phích nước	111	
7	bowl	Bát	140	
8	cap	Mũ lưỡi trai	110	Đồ thời trang
9	T-shirt	Áo phông	106	
10	balo	Ba lô	152	

Bảng 3.1. Chi tiết tập dữ liệu thực nghiệm

Cây thư mục của tập ảnh các sản phẩm thương mại điện tử được sử dụng trong thực nghiệm:



Hình 3.1. Cây thư mục của tập dữ liệu thực nghiệm



Hình 3.2. Các ảnh của mặt hàng "Chảo" trong tập dữ liệu

3.2 Tiến hành thực nghiệm và kết quả

3.2.1 Tiến hành thực nghiệm

Từ mô hình ResNet50 đã trình bày ở Chương II, ta áp dụng mô hình này để nhận dạng các sản phẩm thương mại điện tử của tập dữ liệu hình ảnh thực ở mục trên.

Sử dụng Google Colab để viết chương trình lập trình với ngôn ngữ lập trình python. Với sự tiện lợi của các thư viện, frame work hỗ trợ lập trình hiện nay, ta sử dụng luôn framework Tensorflow để viết chương trình.

- Khởi tạo mô hình với API Sequential().
- Thêm mô hình ResNet50 có sẵn của Tensorflow vào bằng hàm add().
- Sử dụng các trọng số weights được huấn luyện trước từ tập dữ liệu ImageNet. Bỏ đi lớp đầu ra của mô hình có sẵn.
- Thêm một lớp neural network bình thường gồm 10 nút tương ứng với 10 class là 10 loại sản phẩm mà chúng ta cần phân loại. Sử dụng hàm phân loại Softmax.
- Cuối cùng ta có thể xem tổng quan về mô hình cùng tổng số lượng trọng số cần huấn luyện nhờ hàm summary().

The screenshot shows a Google Colab notebook with the following code and output:

```
[41] model = Sequential()
```

```
[42] model.add(ResNet50(include_top = False, pooling = 'avg', weights = 'imagenet'))
      model.add(Dense(10, activation = 'softmax'))
      model.layers[0].trainable = False
```

Downloading data from <https://storage.googleapis.com/tensorflow/keras-applications>
 94773248/94765736 [=====] - 1s 0us/step
 94781440/94765736 [=====] - 1s 0us/step

```
model.summary()
```

Model: "sequential"

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2048)	23587712
dense (Dense)	(None, 10)	20490

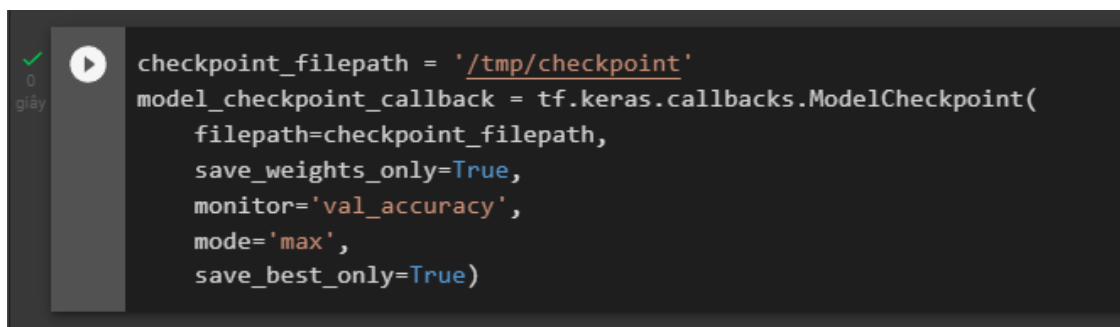
=====
 Total params: 23,608,202
 Trainable params: 20,490
 Non-trainable params: 23,587,712

Hình 3.3. Tổng quan mô hình ResNet50 được sử dụng

Nhìn vào hình 3.3 có thể thấy ngay được output shape, số lượng layer, và tổng số các parmas (các tham số của mô hình). Trong đó thấy rằng các layer của mô hình ResNet50, đã được bỏ đi lớp đầu ra mặc định phân loại 1000 lớp có output shape

(None, 1000), thay vào đó là đầu ra với 10 lớp nhãn out shape(None,10) đã được ta thay đổi thành công.

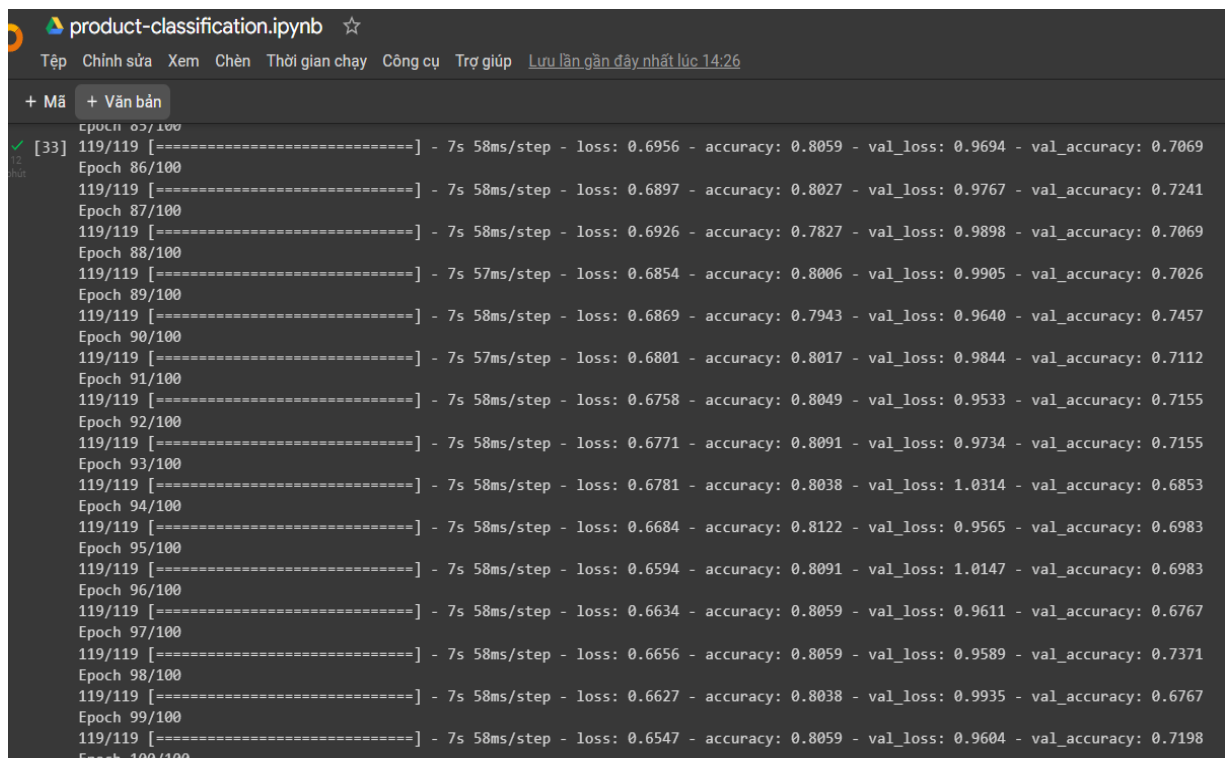
Sau khi có được thiết lập mô hình học sâu để sử dụng, ta viết code để lưu checkpoint. ModelCheckpoint được sử dụng kết hợp với huấn luyện sử dụng model.fit () để lưu mô hình hoặc trọng số (trong tập test) tại một khoảng thời gian nào đó. Ở đây chúng ta điều chỉnh các tham số của hàm để lưu lại các trọng số của epochs tốt nhất trong quá trình đào tạo.



```
checkpoint_filepath = '/tmp/checkpoint'
model_checkpoint_callback = tf.keras.callbacks.ModelCheckpoint(
    filepath=checkpoint_filepath,
    save_weights_only=True,
    monitor='val_accuracy',
    mode='max',
    save_best_only=True)
```

Hình 3.4. Model Checkpoint

Tiếp đến, ta tiến hành đào tạo mô hình ResNet50 này 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.



```

product-classification.ipynb ☆
Tệp  Chỉnh sửa  Xem  Chèn  Thời gian chạy  Công cụ  Trợ giúp  Lưu lần gần đây nhất lúc 14:26

+ Mã  + Văn bản

[33] 119/119 [=====] - 7s 58ms/step - loss: 0.6956 - accuracy: 0.8059 - val_loss: 0.9694 - val_accuracy: 0.7069
Epoch 86/100
119/119 [=====] - 7s 58ms/step - loss: 0.6897 - accuracy: 0.8027 - val_loss: 0.9767 - val_accuracy: 0.7241
Epoch 87/100
119/119 [=====] - 7s 58ms/step - loss: 0.6926 - accuracy: 0.7827 - val_loss: 0.9898 - val_accuracy: 0.7069
Epoch 88/100
119/119 [=====] - 7s 57ms/step - loss: 0.6854 - accuracy: 0.8006 - val_loss: 0.9905 - val_accuracy: 0.7026
Epoch 89/100
119/119 [=====] - 7s 58ms/step - loss: 0.6869 - accuracy: 0.7943 - val_loss: 0.9640 - val_accuracy: 0.7457
Epoch 90/100
119/119 [=====] - 7s 57ms/step - loss: 0.6801 - accuracy: 0.8017 - val_loss: 0.9844 - val_accuracy: 0.7112
Epoch 91/100
119/119 [=====] - 7s 58ms/step - loss: 0.6758 - accuracy: 0.8049 - val_loss: 0.9533 - val_accuracy: 0.7155
Epoch 92/100
119/119 [=====] - 7s 58ms/step - loss: 0.6771 - accuracy: 0.8091 - val_loss: 0.9734 - val_accuracy: 0.7155
Epoch 93/100
119/119 [=====] - 7s 58ms/step - loss: 0.6781 - accuracy: 0.8038 - val_loss: 1.0314 - val_accuracy: 0.6853
Epoch 94/100
119/119 [=====] - 7s 58ms/step - loss: 0.6684 - accuracy: 0.8122 - val_loss: 0.9565 - val_accuracy: 0.6983
Epoch 95/100
119/119 [=====] - 7s 58ms/step - loss: 0.6594 - accuracy: 0.8091 - val_loss: 1.0147 - val_accuracy: 0.6983
Epoch 96/100
119/119 [=====] - 7s 58ms/step - loss: 0.6634 - accuracy: 0.8059 - val_loss: 0.9611 - val_accuracy: 0.6767
Epoch 97/100
119/119 [=====] - 7s 58ms/step - loss: 0.6656 - accuracy: 0.8059 - val_loss: 0.9589 - val_accuracy: 0.7371
Epoch 98/100
119/119 [=====] - 7s 58ms/step - loss: 0.6627 - accuracy: 0.8038 - val_loss: 0.9935 - val_accuracy: 0.6767
Epoch 99/100
119/119 [=====] - 7s 58ms/step - loss: 0.6547 - accuracy: 0.8059 - val_loss: 0.9604 - val_accuracy: 0.7198
Epoch 100/100

```

Hình 3.5. Các epochs trong quá trình đào tạo mô hình

Sau khi đào tạo xong, nhờ phần checkpoint ở bên trên mà ta có thể dễ dàng lưu mô hình cùng các trọng số đã được đào tạo của epochs tốt nhất vào file `product.h5`, lưu lại bằng hàm `save()` có sẵn được Tensorflow cung cấp. Với file `product.h5` chứa mô hình ResNet50 với các trọng số đã được đào tạo qua tập thực nghiệm này, ta có thể đem mô hình đi nhận dạng các sản phẩm thương mại điện tử một cách tiện lợi. Khi cần sử dụng mô hình này, ta chỉ việc sử dụng hàm `load_model()` để load file `product.h5` và sau đó dùng hàm `predict()` như bình thường để nhận dạng các ảnh sản phẩm.

3.2.2. Kết quả thực nghiệm

Với mô hình ResNet50 đã được đào tạo qua tập dữ liệu các hình ảnh sản phẩm thương mại điện tử ở trên. Ta có một số các kết quả dự đoán với các loại sản phẩm như sau:

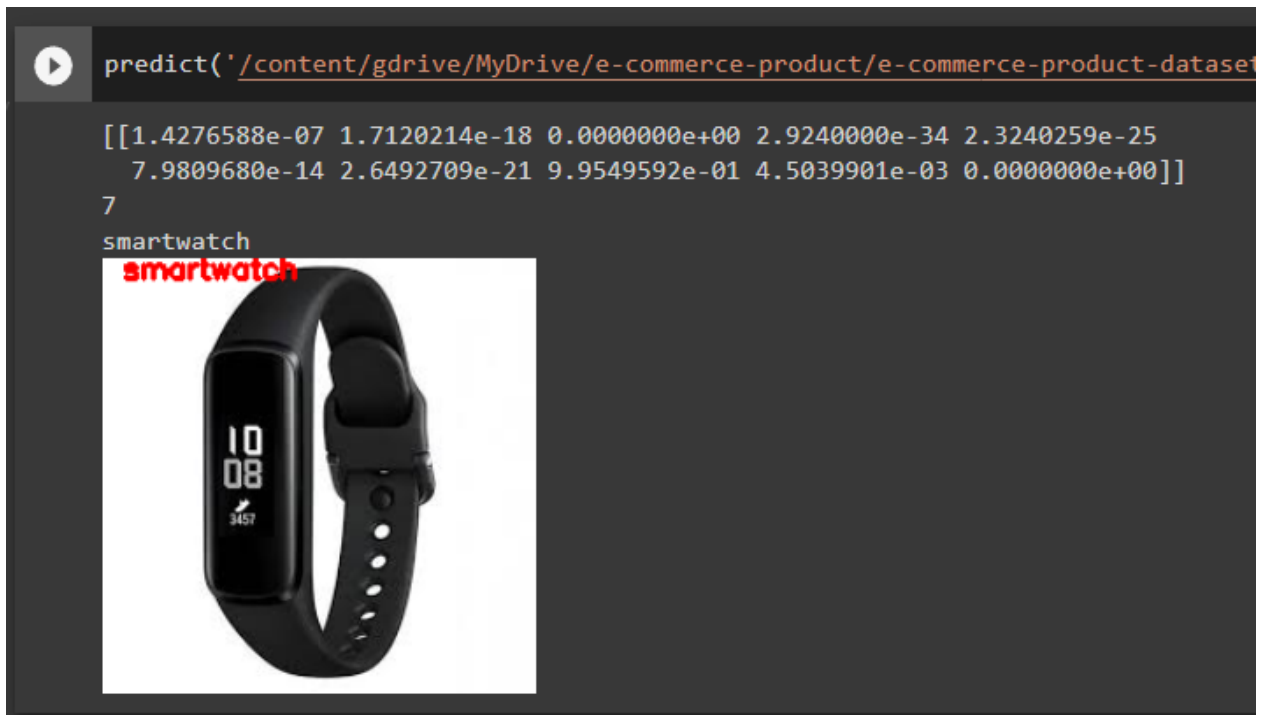
- Kết quả dự đoán của một số ảnh mặt hàng thuộc loại sản phẩm đồ điện tử:
Ảnh mặt hàng smartphone (điện thoại thông minh). Đây là chiếc Iphone13 promax đã được nhận dạng chính xác.



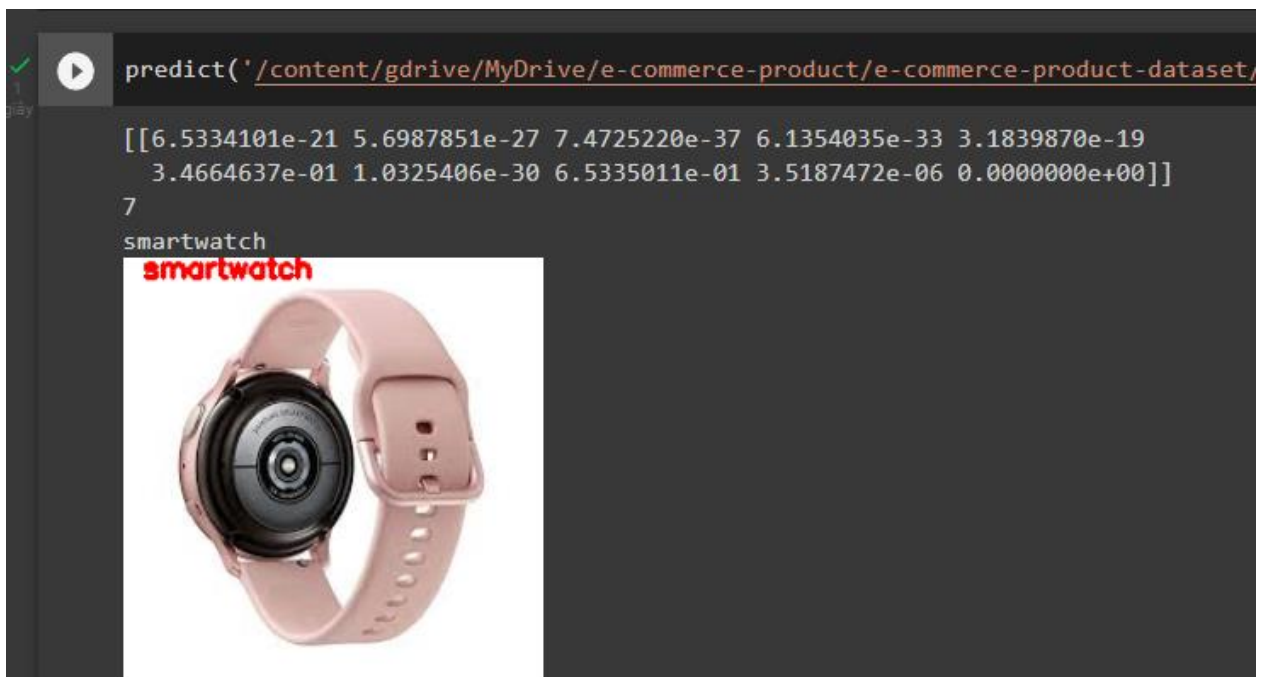
Hình 3.6. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 1

Nhìn vào hình 3.3 ở trên, có thể thấy rằng kết quả dự đoán đầu ra là một mảng gồm 10 số dự đoán xác suất tương ứng với 10 loại sản phẩm. Ta lấy chỉ mục tương ứng của sản phẩm có giá trị xác suất trong mảng là cao nhất. Ở mảng trong hình thì phần tử thứ 6 là số có xác suất cao nhất, ứng với nhãn là smartphone.

Dưới đây là kết quả nhận dạng một số sản phẩm khác cùng thuộc loại mặt hàng điện tử. Hình ảnh nhận dạng 2 chiếc đồng hồ thông minh (smartwatch) khác nhau. Ảnh một chiếc đồng hồ màu đen có góc chụp mặt trước và ảnh một chiếc đồng hồ màu hồng với góc chụp đằng sau. Cả hai sản phẩm đồng hồ thông minh này đều được nhận diện chính xác.



Hình 3.7. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 2

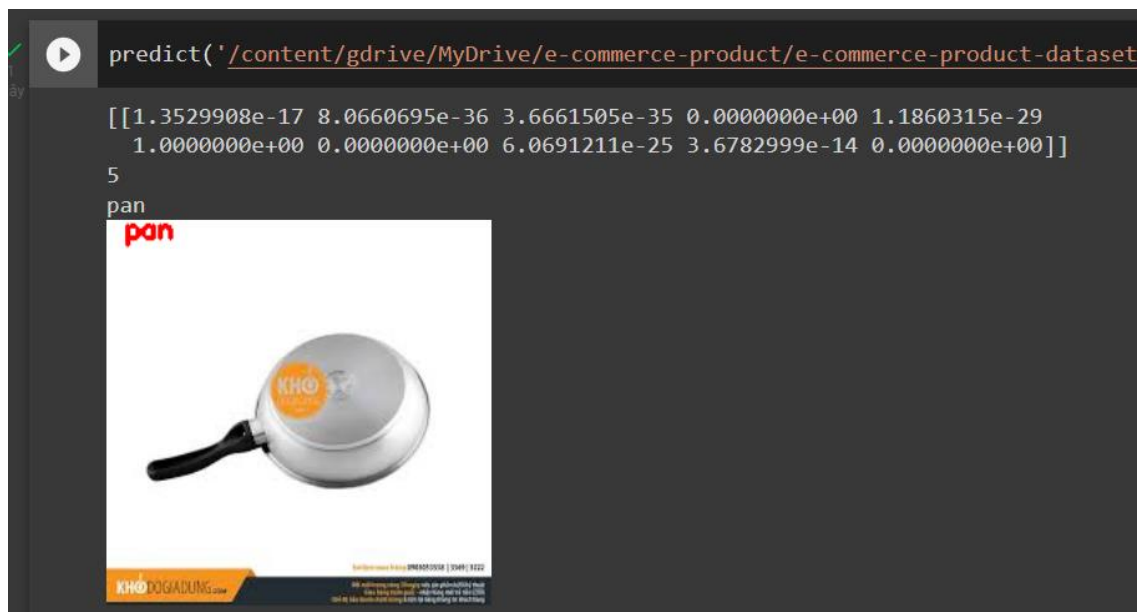


Hình 3.8. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ điện tử 3

- Kết quả dự đoán của một số ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng: Ảnh mặt hàng pan (chảo). Đây là một chiếc chảo nấu ăn, để theo phương ngang, ngửa. Mô hình cho kết quả nhận diện chính xác.



Hình 3.9. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 1
Với ảnh của chiếc chảo bị lật úp như bên dưới, mô hình vẫn cho kết quả chính xác.



Hình 3.10. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 2

Ảnh của một loại mặt hàng khác là quạt cây (treefan) được nhận dạng bởi mô hình:



Hình 3.11. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm đồ gia dụng 3

- Kết quả dự đoán của một số hình ảnh mặt hàng thuộc loại sản phẩm đồ thời trang:

Ảnh của một loại mặt hàng thời trang là mũ lưỡi trai (cap) được nhận dạng bởi mô hình:



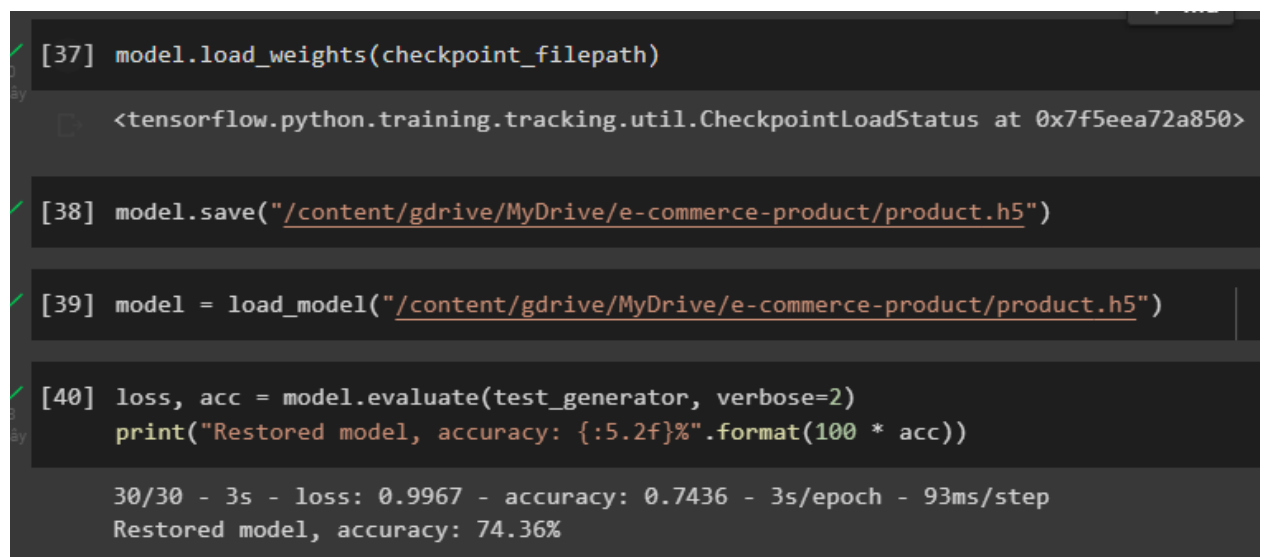
Hình 3.12. Kết quả nhận dạng ảnh mặt hàng thuộc loại sản phẩm thời trang

3.3 Đánh giá kết quả

3.3.1. Đánh giá bằng độ Accuracy

Có rất nhiều phương pháp để đánh giá kết quả của mô hình nhưng phương pháp được sử dụng nhiều và phổ biến rộng rãi nhất đó là tính độ chính xác 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.

Sử dụng hàm evaluate() của framework Tensorflow để đánh giá kết quả của model trên tập test. Ta thấy độ chính xác của mô hình trên tập test là 74.36%. Đây được xem là một kết quả khá tốt cho một mô hình dự đoán với số lượng lớp cũng nhiều và lượng dữ liệu ít như tập dữ liệu ảnh thực nghiệm ở trên.



```
[37] model.load_weights(checkpoint_filepath)
<tensorflow.python.training.tracking.util.CheckpointLoadStatus at 0x7f5eea72a850>

[38] model.save("/content/gdrive/MyDrive/e-commerce-product/product.h5")

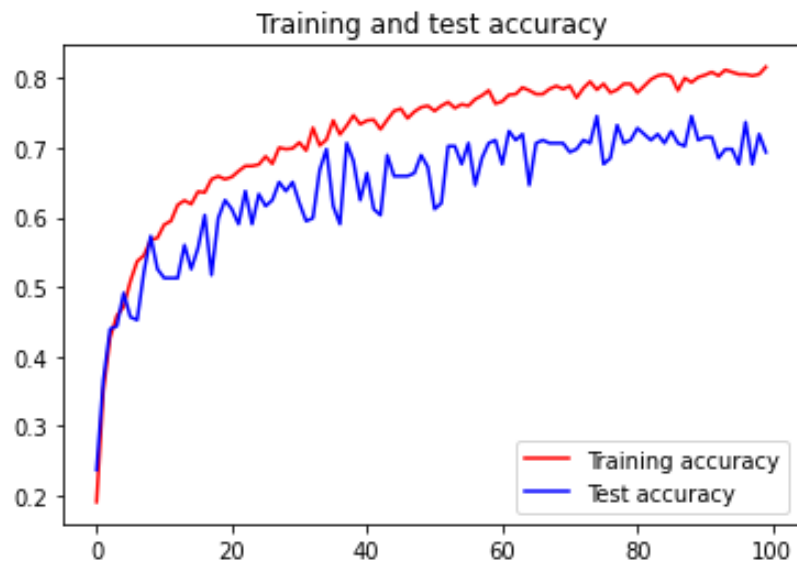
[39] model = load_model("/content/gdrive/MyDrive/e-commerce-product/product.h5")

[40] loss, acc = model.evaluate(test_generator, verbose=2)
print("Restored model, accuracy: {:.2f}%".format(100 * acc))

30/30 - 3s - loss: 0.9967 - accuracy: 0.7436 - 3s/epoch - 93ms/step
Restored model, accuracy: 74.36%
```

Hình 3.13. Độ chính xác Accuracy trên tập kiểm thử

Nhìn vào hình 3.5 của phần trước, ta thấy sau mỗi epochs thì thường độ chính xác accuracy của mô hình trên cả tập train và tập test đều có xu hướng tăng dần. Điều này có nghĩa là sau mỗi lần học, mô hình được đào tạo chạy qua một lần tập dữ liệu thì nó sẽ chính xác hơn, mô hình trở lên thông minh hơn. Nhìn vào hình 38 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.



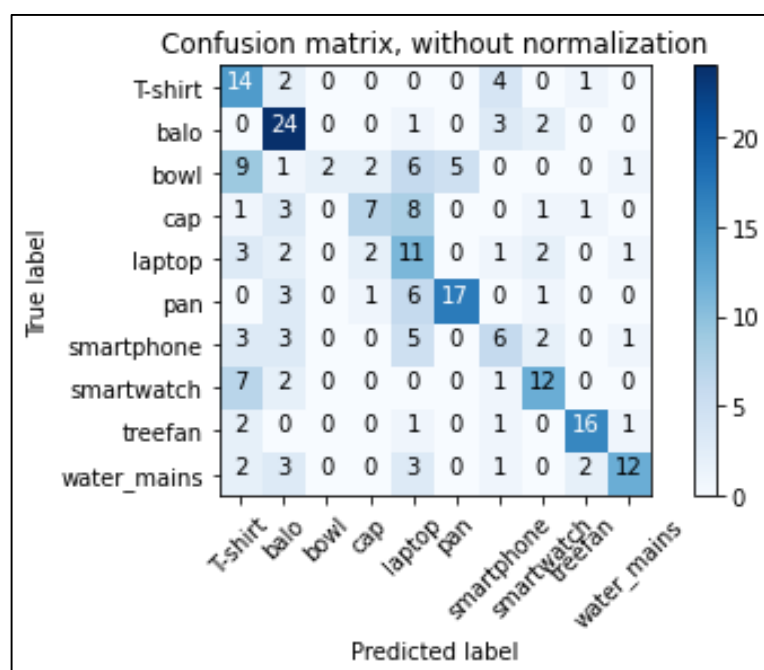
Hình 3.14. Biểu đồ đường độ chính xác accuracy qua các epochs

3.3.2. Đánh giá bằng *Confusion matrix*

Cách tính sử dụng accuracy như ở trên chỉ cho chúng ta biết được bao nhiêu phần trăm lượng dữ liệu được phân loại đúng mà không chỉ ra được cụ thể mỗi loại được phân loại như thế nào, lớp nào được phân loại đúng nhiều nhất, và dữ liệu thuộc lớp nào thường bị phân loại nhầm vào lớp khác. Để có thể đánh giá được các giá trị này, chúng ta sử dụng một ma trận được gọi là confusion matrix.

Về cơ bản, confusion matrix thể hiện có bao nhiêu điểm dữ liệu thực sự thuộc vào một class, và được dự đoán là rơi vào một class. Nó là một ma trận vuông với kích thước mỗi chiều bằng số lượng lớp dữ liệu. Giá trị tại hàng thứ i , cột thứ j là số lượng điểm lẽ ra thuộc vào class i nhưng lại được dự đoán là thuộc vào class j . Chúng ta có thể suy ra ngay rằng tổng các phần tử trong toàn ma trận này chính là số điểm trong tập kiểm thử. Các phần tử trên đường chéo của ma trận là số điểm được phân loại đúng của mỗi lớp dữ liệu. Từ đây có thể suy ra accuracy chính bằng tổng các phần tử trên đường chéo chia cho tổng các phần tử của toàn ma trận.

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



Hình 3.15. 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. Lớp 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

Precision được định nghĩa là tỉ lệ số điểm Positive mô hình dự đoán đúng trên tổng số điểm mô hình dự đoán là Positive.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (8)$$

Recall được định nghĩa là tỉ lệ số điểm Positive mô hình dự đoán đúng trên tổng số điểm thật sự là Positive (hay tổng số điểm được gán nhãn là Positive ban đầu).

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (9)$$

Precision càng cao, tức là số điểm mô hình dự đoán là positive đều là positive càng nhiều. Precision = 1, tức là tất cả số điểm mô hình dự đoán là Positive đều đúng, hay không có điểm nào có nhãn là Negative mà mô hình dự đoán nhầm là Positive.

Recall càng cao, tức là số điểm là positive bị bỏ sót càng ít. Recall = 1, tức là tất cả số điểm có nhãn là Positive đều được mô hình nhận ra.

Tuy nhiên, chỉ có Precision hay chỉ có Recall thì không đánh giá được chất lượng mô hình.

- Chỉ dùng Precision, mô hình chỉ đưa ra dự đoán cho một điểm mà nó chắc chắn nhất. Khi đó Precision = 1, tuy nhiên ta không thể nói là mô hình này tốt.
- Chỉ dùng Recall, nếu mô hình dự đoán tất cả các điểm đều là positive. Khi đó Recall = 1, tuy nhiên ta cũng không thể nói đây là mô hình tốt.

Khi đó F1-score được sử dụng. F1-score là trung bình điều hòa (harmonic mean) của Precision và Recall (giả sử hai đại lượng này khác 0). F1-score được tính theo công thức:

$$\frac{2}{\text{F1}} = \frac{1}{\text{Precision}} + \frac{1}{\text{Recall}} \quad (10)$$

Đối với các bài toán phân lớp nhiều lớp như nhận dạng sản phẩm thương mại điện tử, ta lần lượt xem một lớp là Positive, các lớp còn lại là Negative. Khi đó, ta có hai cách tính F1-score: Macro F1-score và Micro F1-score.

Macro F1-score được tính bằng công thức với Macro-Precision và Macro-Recall là trung bình cộng của Precision và Recall theo các lớp:

$$\frac{2}{\text{Macro-F1}} = \frac{1}{\text{Macro-Precision}} + \frac{1}{\text{Macro-Recall}} \quad (11)$$

Micro F1-score được tính bằng công thức với Micro-Precision là tỉ lệ tổng số điểm của toàn bộ các điểm Positive thuộc các lớp trên tổng số điểm mô hình dự đoán là Positive thuộc các lớp. Micro-Recall được tính bằng cách tương tự.

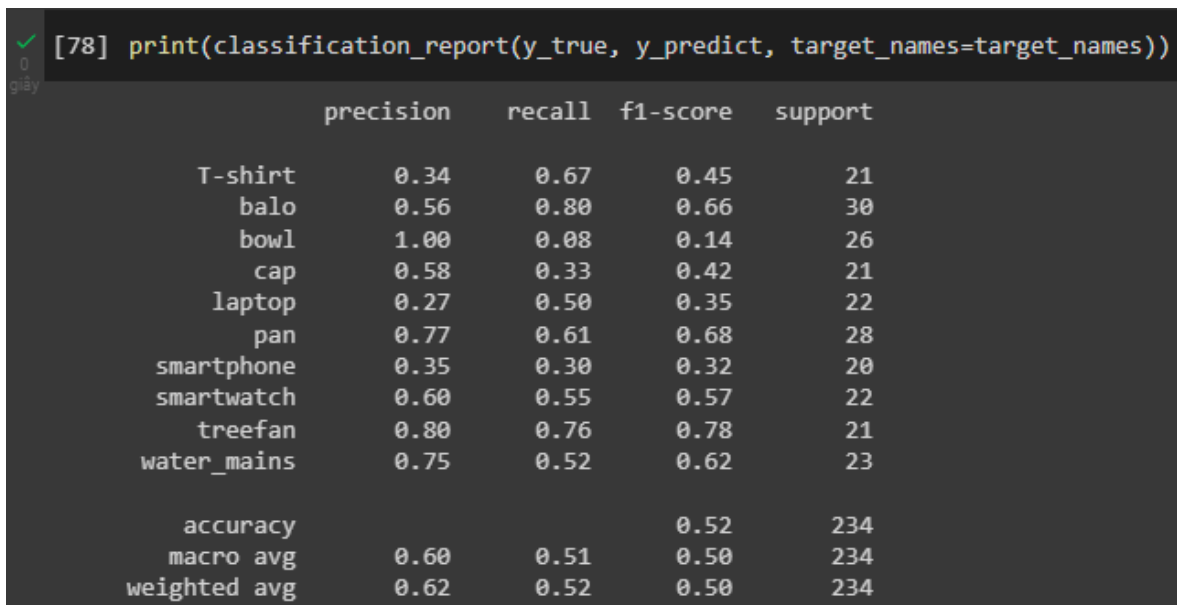
$$\frac{2}{\text{Micro-F1}} = \frac{1}{\text{Micro-Precision}} + \frac{1}{\text{Micro-Recall}} \quad (12)$$

$$\text{Micro-Precision} = \frac{\sum_c^C \text{TP}_c}{(\text{TP}_c + \text{FP}_c)} \quad (13)$$

$$\text{Micro-Recall} = \frac{\sum_c^C \text{TP}_c}{(\text{TP}_c + \text{FN}_c)} \quad (14)$$

Với TP_c , FP_c , FN_c lần lượt là TP, FP, FN của class c trong tất cả C class.

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 3.16. 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

Với sự phát triển hiện nay của công nghệ, sự phổ biến của mạng internet thì các trang web thật sự rất tiện lợi. Ta có thể truy cập trang web và sử dụng tiện ích của nó một cách đơn giản chỉ cần một thiết bị có trình duyệt web và có kết nối internet bất kể nó là máy tính để bàn, laptop, smartphone... Vậy nên xây dựng một hệ thống nhận dạng sản phẩm dưới dạng web là điều thật sự tiện ích và hiệu quả cho người dùng.

Từ mô hình ResNet50 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

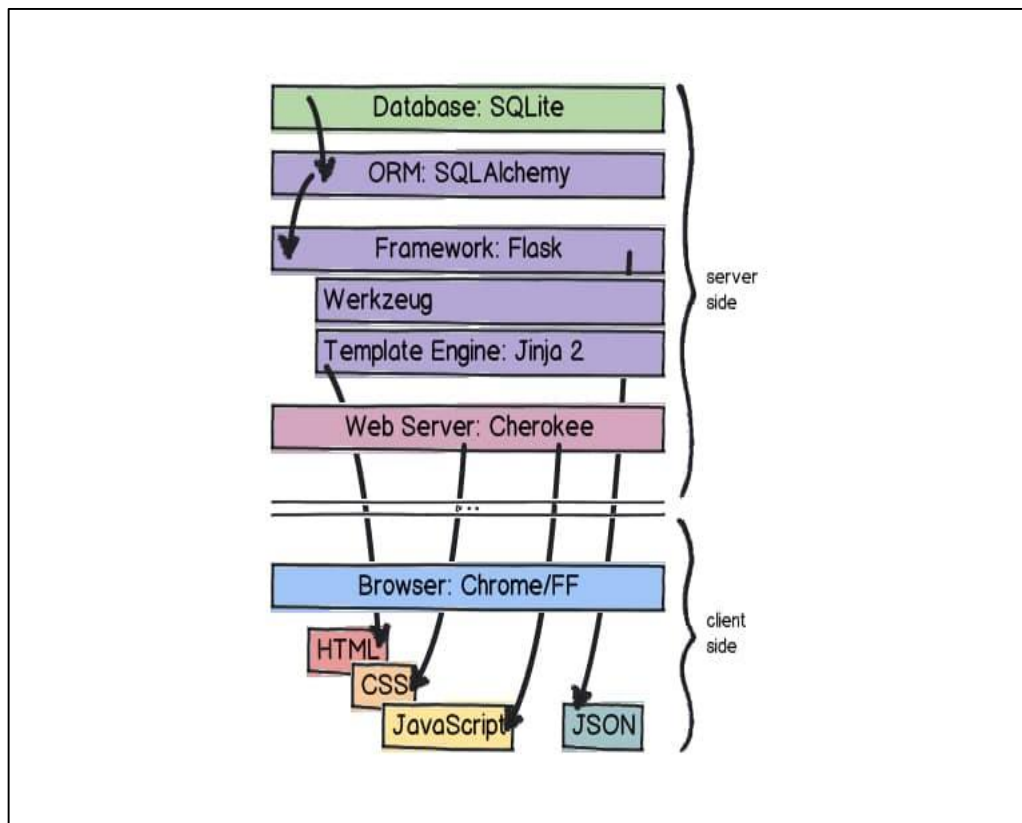
Flask là một Web Framework rất nhẹ của Python, dễ dàng giúp người mới bắt đầu học Python có thể tạo ra website nhỏ. Flask cũng dễ mở rộng để xây dựng các ứng dụng web phức tạp. Flask có nền tảng là Werkzeug WSGI và Jinja2 và nó đã trở thành một trong những Web Framework phổ biến nhất của Python.

Giao diện cổng máy chủ web (Web Server Gateway Interface, WSGI) đã được sử dụng như một tiêu chuẩn để phát triển ứng dụng web Python. WSGI là đặc điểm kỹ thuật của giao diện chung giữa máy chủ web và ứng dụng web.

Werkzeug là một bộ công cụ WSGI thực hiện các yêu cầu, đối tượng phản hồi và các chức năng tiện ích. Điều này cho phép một khung web được xây dựng trên đó. Khung công tác Flask sử dụng Werkzeug làm một trong những cơ sở của nó.

Jinja2 là một công cụ mẫu phổ biến cho Python. một hệ thống mẫu web kết hợp một mẫu với một nguồn dữ liệu cụ thể để hiển thị một trang web động.

Flask Framework là một bộ lưu trữ giúp các lập trình viên tạo ra các trang web dễ dàng hơn, có thể mở rộng, hiệu quả và có thể bảo trì bằng cách cung cấp code hoặc tiện ích mở rộng có thể sử dụng lại cho các nhiệm vụ phổ biến.



Hình 3.17. Cách Flask Framework hoạt động

Flask thường được coi là một microframework. Nó được thiết kế để giữ cho cốt lõi của ứng dụng đơn giản và có thể mở rộng. Thay vì một lớp trừu tượng để hỗ trợ cơ sở dữ liệu, Flask hỗ trợ các phần mở rộng để thêm các khả năng như vậy vào ứng dụng.

Như đã nêu trước đó, Flask được phân loại là Web Framework siêu nhỏ, nhẹ. Thông thường, một framework vi mô là một framework tối giản hoặc không phụ thuộc vào thư viện bên ngoài. Nó có rất nhiều ưu điểm: tốc độ, hỗ trợ cho NoQuery, độ phức tạp tối thiểu, tối giản tuyệt đối, không có ORM, dễ dàng kết nối với tiện ích mở rộng, trình gỡ lỗi được nhúng trong trình duyệt, mã ngắn và đơn giản trong số các framework Python khác...

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.

```

index.html X
templates > index.html > ...
33         method="POST"
34         enctype="multipart/form-data"
35     <fieldset>
36         <div class="form-horizontal">
37             <div class="form-group">
38                 <div class="row">
39                     <div class="col-md-10">
40                         <div class="input-group">
41                             <input type="hidden" id="filename" name="filename" value="">
42                             <input type="file" id="uploadedFile" name="query_img" class="form-control form-control-sm">
43                             <div class="input-group-btn">
44                                 <input type="submit" value="Nhận dạng" class="rounded-0 btn btn-primary">
45                             </div>
46                         </div>
47                     </div>
48                 </div>
49             </div>
50         </div>
51     </fieldset>
52 </form>
53 </div>
54 <span>
55     {% if query_path %}
56     <h5>ẢNH SẢN PHẨM: </h5>
57     
58     <h5>KẾT QUẢ: </h5>
59     <h6>
60     Tên sản phẩm:<figcaption>{{ pred_label_product }}</figcaption>
61     </h6>
62     <h7>
63     Loại sản phẩm:<figcaption>{{ pred_category_product }}</figcaption>
64     </h7>
65     {% endif %}
66 </div>
67 </body>
68 </html>
69

```

Hình 3.18. File giao diện index.html của Web nhận dạng sản phẩm TMĐT

```

main.py X
main.py
33     category_product = "Đồ thời trang"
34
35     return predicted_product,category_product
36
37 #build web Flask
38 app = Flask(__name__)
39 @app.route('/', methods=['GET', 'POST'])
40
41 def index():
42     if request.method == 'POST':
43         file = request.files['query_img']
44         # Save query image
45         img = Image.open(file) # PIL image
46         uploaded_img_path = "static/uploaded/"+ file.filename
47         img.save(uploaded_img_path)
48         pred_label_product, pred_category_product = predict(uploaded_img_path)
49         return render_template('index.html',
50                               query_path=uploaded_img_path,
51                               pred_label_product = pred_label_product,
52                               pred_category_product= pred_category_product)
53     else:
54         return render_template('index.html')
55
56 if __name__ == "__main__":
57     app.run(host='127.0.0.1', port=8080, debug=True)

```

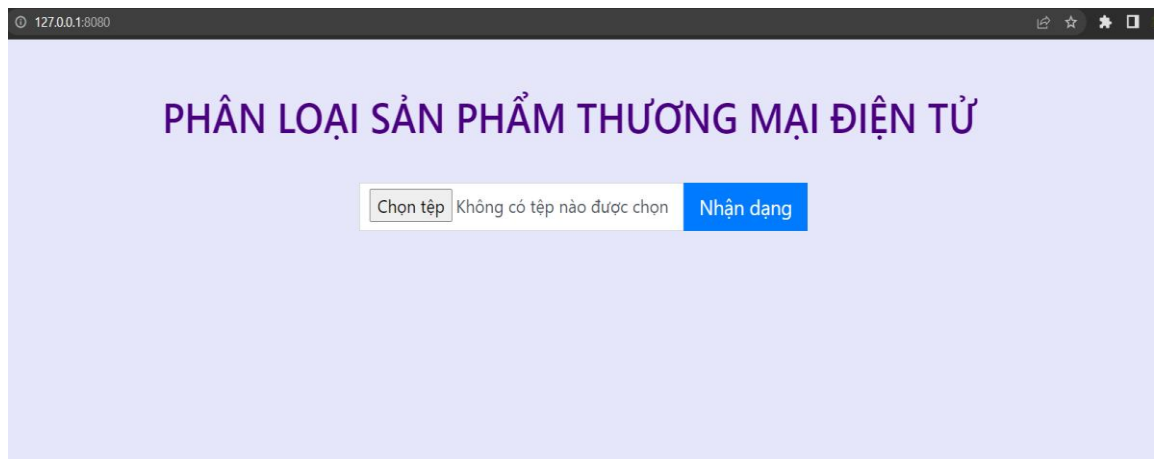
Hình 3.19. API RESTful GET và POST của Flask trong Web nhận dạng sản phẩm thương mại điện tử

3.4.4 Trang web nhận dạng sản phẩm thương mại điện tử

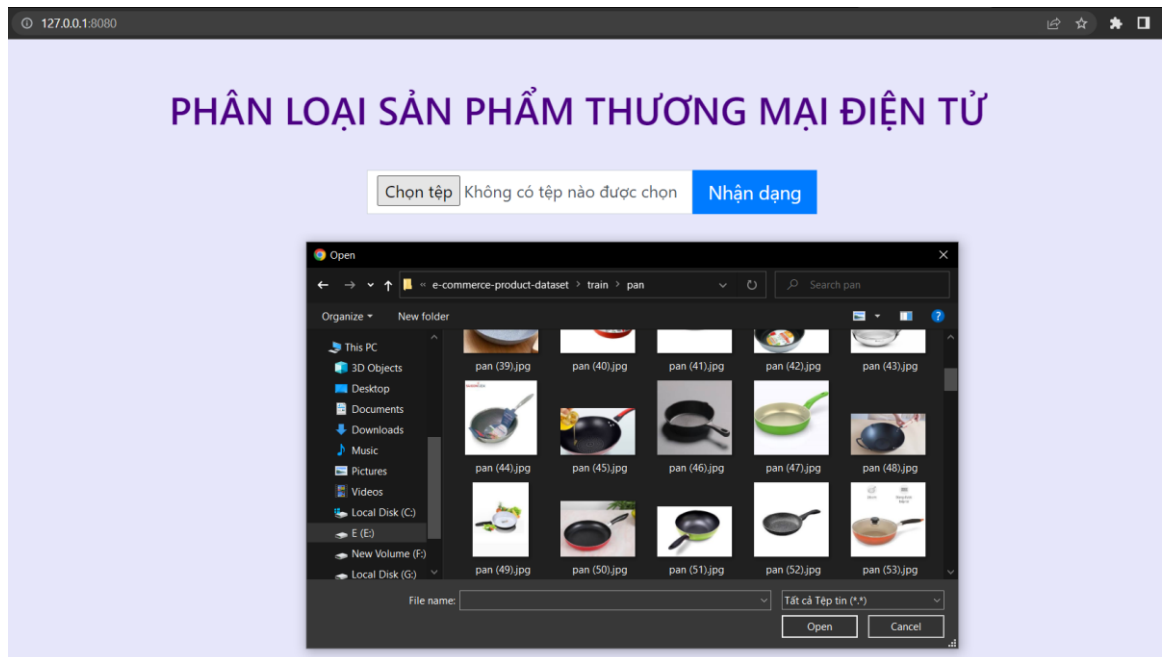
Sau khi có giao diện hoàn chỉnh, kết hợp cùng với mô hình ResNet50 đã được đào tạo ở trên. Ta được website nhận dạng sản phẩm thương mại điện tử có chức năng nhận diện phân loại được ảnh sản phẩm thương mại điện tử mà người dùng tải ảnh lên. Giao diện trang web có một nút “Chọn tệp” để upload file ảnh và một nút “Nhận dạng” để nhận dạng sản phẩm. Khi người dùng muốn nhận dạng hình ảnh:

- Đầu tiên là click nút “Chọn tệp” để tải ảnh truy vấn.
- Hệ thống sẽ mở cửa sổ chọn file ảnh trên thư mục máy tính.
- Người dùng chọn ảnh trong các thư mục máy tính của họ và tải lên.
- Sau khi ảnh truy vấn đã được tải lên thành công, người dùng click vào nút “Nhận dạng”.
- Hệ thống thực hiện phân loại hình ảnh sản phẩm. Trả về kết quả tên sản phẩm cùng loại sản phẩm thương mại điện tử.

Dưới đây là giao diện bắt đầu 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.



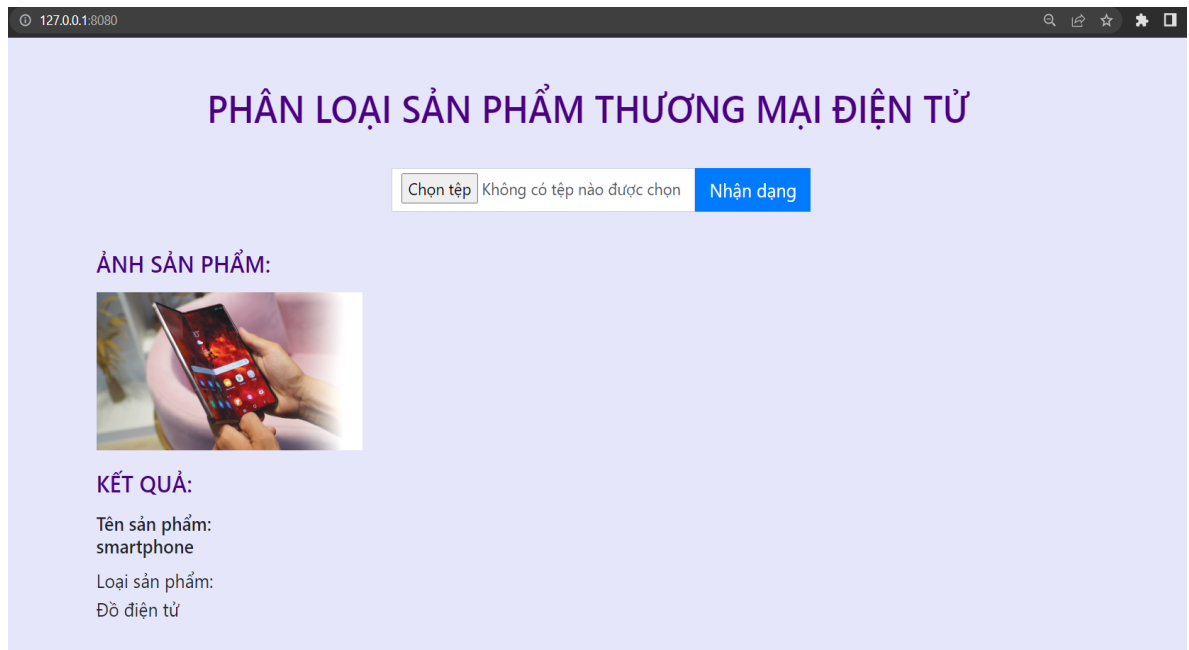
Hình 3.20. Trang giao diện web nhận dạng sản phẩm TMĐT



Hình 3.21. Cửa sổ chọn file ảnh truy vấn trên thư mục máy tính



Hình 3.22. Web hiển thị ảnh sản phẩm và kết quả nhận dạng



Hình 3.23. Một kết quả nhận dạng sản phẩm khác của web

3.5 Kết luận chương III

Trong chương này, luận văn đã tiến hành thực nghiệm mô hình ResNet50 với tập dữ liệu hình ảnh các sản phẩm thương mại điện tử. Tập dữ liệu ảnh sản phẩm thương mại điện tử với 3 loại sản phẩm: đồ điện tử, đồ gia dụng, đồ thời trang. Tiến hành đào tạo cho mô hình ResNet50 sử dụng framework Tensorflow. Đánh giá mô hình dựa trên độ chính xác Accuracy. Mô hình ResNet50 sau khi được đào tạo cho độ chính xác 74.36% trên tập kiểm thử.

Luận văn cũng đã trình bày về framework để lập trình web Flask. Đồng thời cũng ứng dụng framework này 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.

TÀI LIỆU THAM KHẢO

Tiếng Việt:

- [1] Vũ Hữu Tiếp, “K-means Clustering”, Machine learning cơ bản:
<https://machinelearningcoban.com/2017/01/01/kmeans/>, 1 Jan 2107.
- [2] Nguyễn Thanh Tuấn, “Sách Deep Learning Cơ Bản”, 8 Aug 2020.
- [3] Phạm Đình Khánh, “Các kiến trúc CNN hiện đại”,
<https://phamdingkhanh.github.io/2020/05/31/CNNHistory.html>, 31 May 2020

Tiếng Anh:

- [1] Wesley Chai “Definition E-Commerce”, Techtarget,
<https://www.techtarget.com/searchcio/definition/e-commerce>, December, 2020
- [2] Michael Keenan “Global Ecommerce Explained: Stats and Trends to Watch in 2022”, Shopify Plus, <https://www.shopify.com/enterprise/global-ecommerce-statistics>, Feb 16, 2022
- [3] Kirti Solanki, “E-Commerce Definition, Types, Features, Advantages & Disadvantages”, Top4u, <https://www.toppers4u.com/2021/01/e-commerce-types-features-advantages.html>, January 20, 2021
- [4] Aakash Kaushik, Opengenius , “Understanding ResNet50 architecture”,
<https://iq.opengenus.org/resnet50-architecture/>
- [5] Nachi Hebbar, Neural Networks , “Transfer Learning With Keras(Resnet-50)”,
<https://chroniclesofai.com/transfer-learning-with-keras-resnet-50/>, Jul 16, 2021