HỆ THỐNG KHUYẾN NGHỊ SẢN PHẨM DỰA TRÊN HÌNH ẢNH VỚI MẠNG NEURAL TÍCH CHẬP

Lê Thanh Phong[1], Nguyễn Văn Phúc Nhân[2] and Vũ Minh Quang[3]

Industrial University of Ho Chi Minh City

Computer Vision Course
Ngày 15 tháng 10 năm 2022

Tổng quan

1. Giới thiệu
2. Phương pháp tiếp cận
   - A. CNN Layers
   - B. Phân loại
   - C. Khuyến nghị
3. Dữ liệu và đặc trưng
4. Thử nghiệm
   - A. Xử lý dữ liệu
   - B. Đánh giá
   - C. Phân loại
   - D. Khuyến nghị
5. Kết luận

Nội dung trình bày

Giới thiệu

Cuộc khủng hoảng của dịch COVID-19 đã mang đến những thách thức nhưng đồng thời cũng là cơ hội để các nhà bán lẻ thay đổi để bắt kịp sự phát triển của thương mại điện tử. Trong bối cảnh đó, thương mại điện tử đang bùng nổ ở nhiều khu vực trên toàn cầu.
Tuy nhiên, điều này đã làm hạn chế khả năng tiếp cận kịp thời về như cầu mua sắm các sản phẩm mà khách hàng quan tâm do tình trạng quá tải thông tin đã xảy ra với khách hàng.
Trong những năm gần đây, với sự phát triển nhanh chóng của mạng neural. Giờ đây, chúng ta có thể thay đổi cách tìm các sản phẩm thông qua mô tả sản phẩm hay tìm kiếm bằng tên sản phẩm bằng hình ảnh của sản phẩm.

Giới thiệu

Phương pháp tìm kiếm bằng hình ảnh đã được áp dụng rất nhiều, nhưng riêng với lĩnh vực thương mại điện tử nói chung và mua sắm trực tuyến nói riêng thì vẫn chưa được ứng dụng rộng rãi.
Dựa trên ý tưởng này, ở đây chúng tôi xây dựng một hệ thống khuyến nghị thông minh từ việc lấy hình ảnh của các đối tượng sản phẩm làm đầu vào thay vì mô tả văn bản như cách truyền thống.
Đầu vào cho thuật toán là hình ảnh của bất kỳ đối tượng sản phẩm nào mà khách hàng muốn mua. Sau đó, sử dụng Convolution Neural Network (CNN) để phân loại.

Giới thiệu

Input: Hình ảnh của 1 sản phẩm
Output: K hình ảnh sản phẩm tương tự hình ảnh đầu vào

Giới thiệu

Cụ thể là, sử dụng Convolution Neural Network để phân loại hình ảnh xem nó thuộc đối tượng nàp mà sản phẩm này có thể thuộc về và sử dụng vector đầu vào của lớp được kết nối đầy đủ cuối cùng dưới dạng vector đặc trưng để tìm kiếm các sản phẩm gần nhất (có liên quan) trong bộ dữ liệu. Thông qua hai bước:

1. **Phân loại**
2. **Khuyến Nghị**

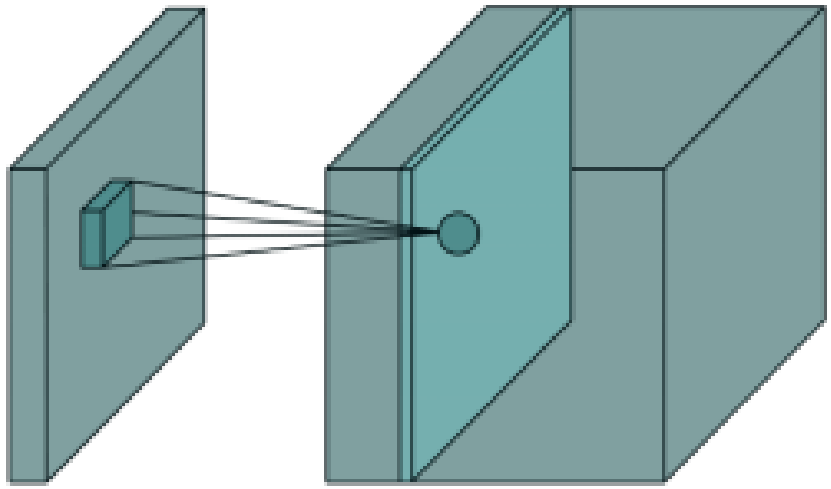
Nội dung trình bày

A. CNN Layers

Hai vấn đề lớn cần giải quyết
Đầu tiên, xác định danh mục mà một hình ảnh nhất định thuộc về.
Thứ hai, tìm và giới thiệu các sản phẩm tương tự nhất theo hình ảnh đã cho trước.

A. CNN Layers

Bước quan trọng nhất của CNN là **lớp tích chập (Conv)**.

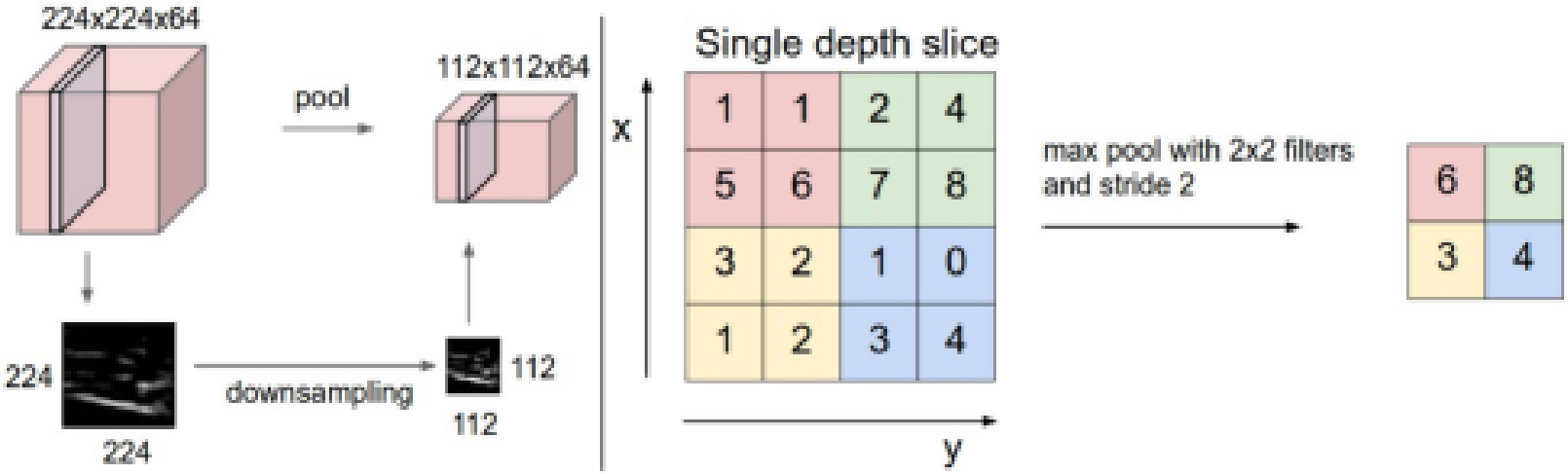


Hình 1. Lớp tích chập

A. CNN Layers

Pooling Layer

Tương tự như các lớp chập, ngoại trừ rằng nó sẽ sử dụng phương thức không tham số để biến đổi nhỏ hình chữ nhật thành một số.



Hình 2. Pooling layers

B. Phân loại

Chúng tôi xây dựng mô hình **AlexNet**, mô hình **VGG-16**, mô hình **VGG-19** và mô hình **ResNet50** cho nhiệm vụ phân loại và so sánh chúng với mô hình SVM làm mô hình cơ sở.

B. Phân loại

Support Vector Machine: Mô hình này về cơ bản là một lớp được kết nối đầy đủ. Chúng tôi sử dụng độ lỗi **Multi-class Support Vector Machine (SVM)** cộng với **L2 Norm** để làm hàm mất mát.

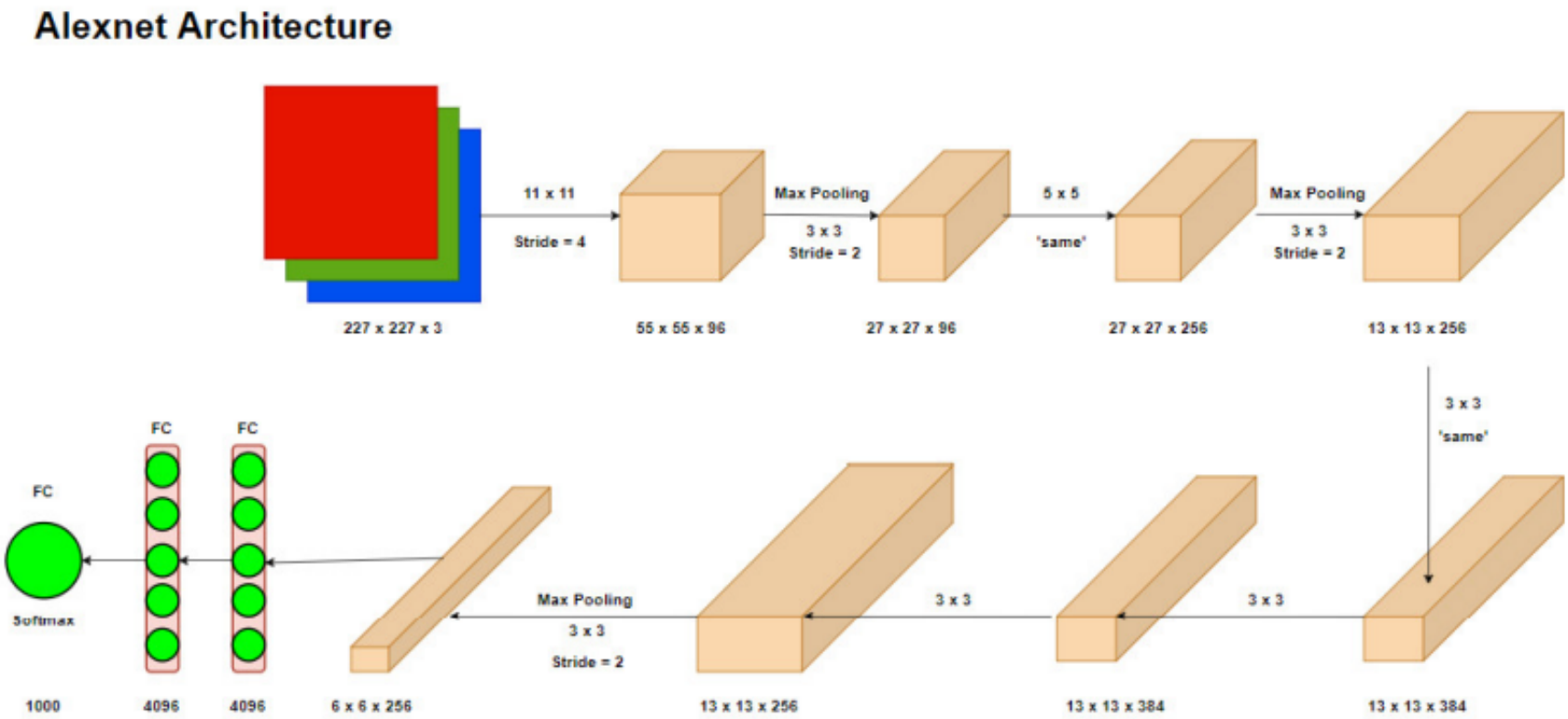
$$s = Wx_i + b \ (1)$$

Trong đó $W \in \mathbb{R}^{n \times d}$ là ma trận trọng số và $b \in \mathbb{R}^n$ là trọng số.
Tổn thất của **SVM** được tính bởi công thức.

$$L_{SVM}\left(W, b; x_i\right) = \sum_{j \neq y_i} max\left(0, s_j - s_{yi} + 1\right) \ (2)$$

Trong đó y_i là nhãn đúng của lớp thực sự.

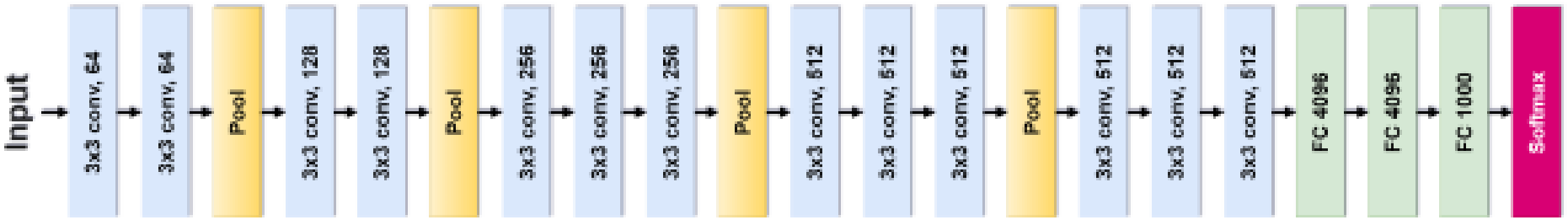
B. Phân loại



Hình 3. Mô hình AlexNet

B. Phân loại

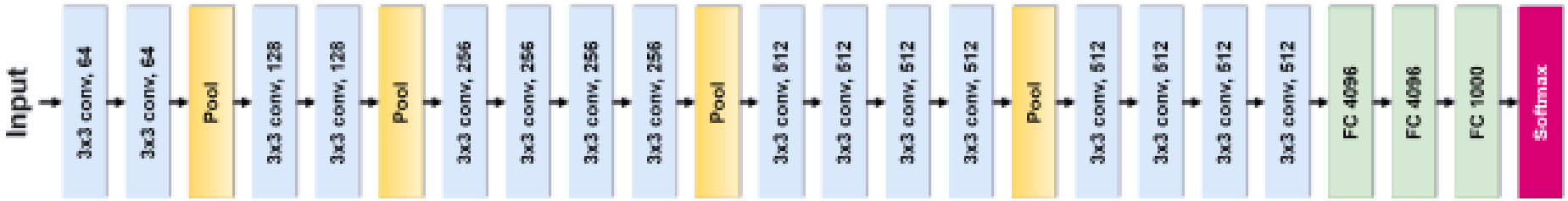
VGG 16 Architecture



Hình 4. Mô hình VGG 16

B. Phân loại

VGG 19 Architecture



Hình 5. Mô hình VGG 19

B. Phân loại

VGG 19 Architecture Ngoài ra chúng tôi còn thêm cũng thêm các lớp chuẩn hóa hàng loạt (batch normalization) sau các hàm kích hoạt để tăng tốc độ huấn luyện cho mô hình và trách trường hợp over-fitting. Và công thức được định nghĩa như sau

$$\mu_i = \frac{1}{m}\sum_{j=1}^{m} x_{ij}$$

$$\sigma_i^2 = \frac{1}{m}\sum_{j=1}^{m}(x_{ij} - \mu_i)^2$$

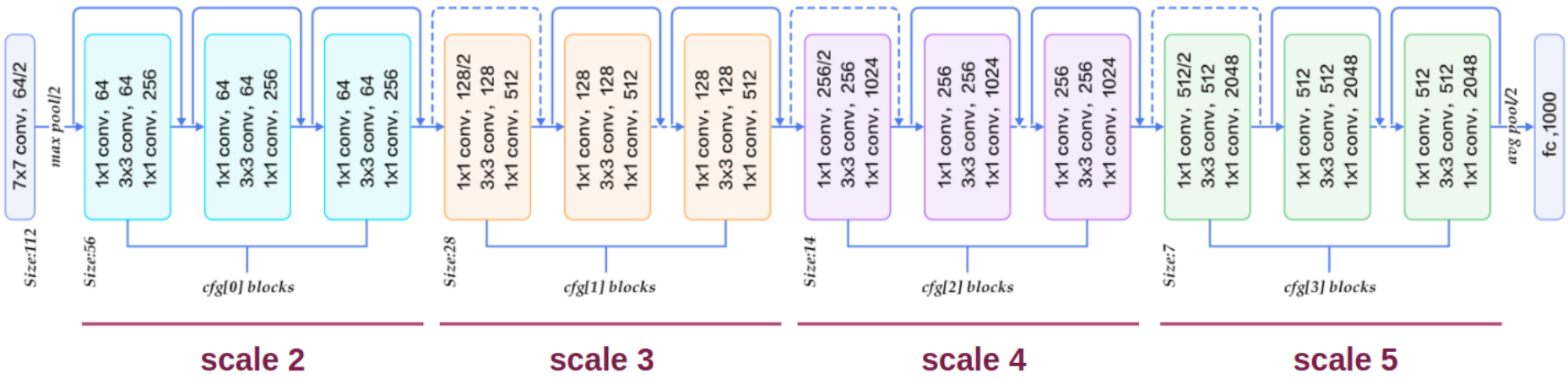
$$\hat{x}_{ij} = \frac{x_{ij} - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}}$$

B. Phân loại

VGG 19 Architecture Batch normalization là một kỹ thuật để đào tạo mạng nơ ron sâu, chuẩn hóa các đầu vào thành một layer cho mỗi mini-batch. Điều này có tác dụng ổn định quá trình học tập và giảm đáng kể số lượng epoch đào tạo cần thiết để đào tạo mạng sâu.
Mục tiêu của phương pháp này chính là việc muốn chuẩn hóa các feature (đầu ra của mỗi layer sau khi đi qua các activation) về trạng thái zero-mean với độ lệch chuẩn 1

B. Phân loại

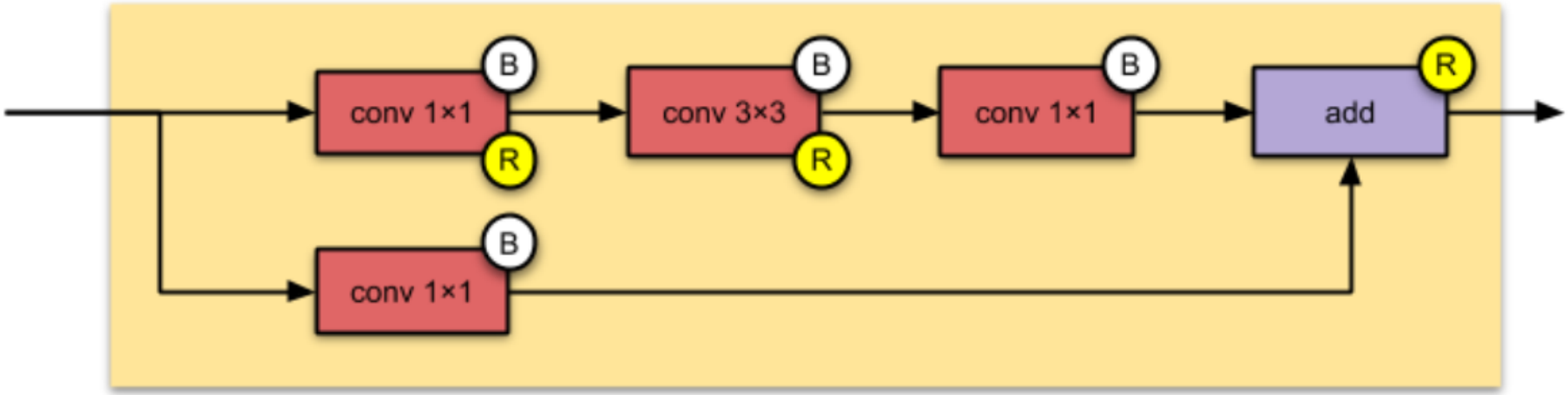
ResNet-50 Architecture



Hình 6. Mô hình ResNet 50

B. Phân loại

ResNet-50 Architecture



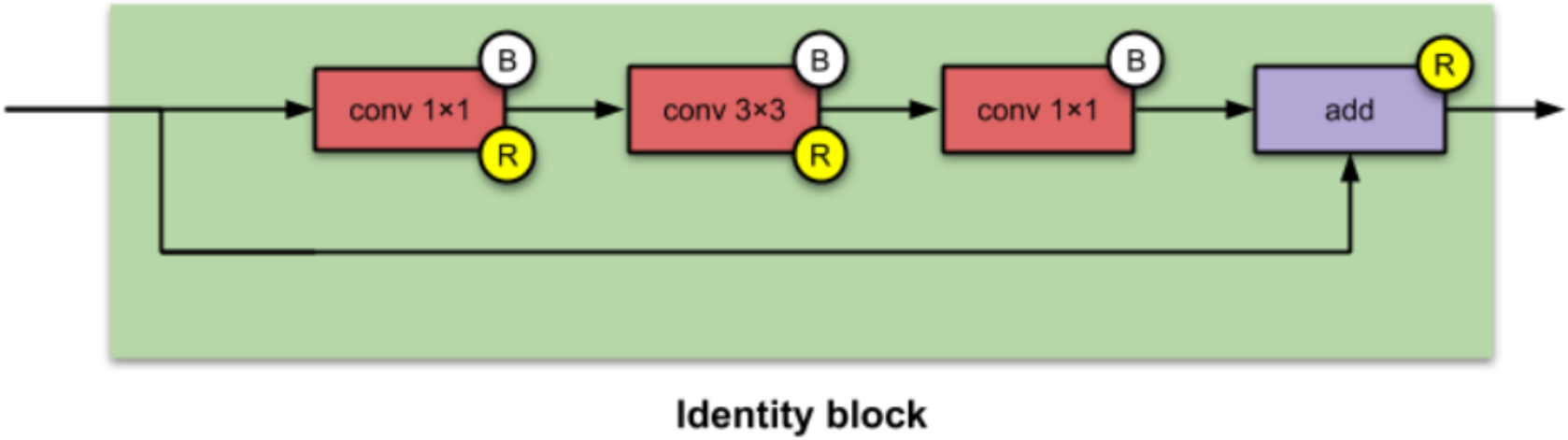
Conv block

Khối tích chập (Convolutional Bock)

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.

B. Phân loại

ResNet-50 Architecture



Identity block

Khối xác định (Identity block)

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.

C. Khuyến nghị

Sử dụng lớp **Fully Connected** cuối cùng trong mô hình phân loại dưới dạng vectơ đặc trưng của hình ảnh.

Đối với bất kỳ hình ảnh nào trong tập dữ liệu, sẽ có một vector đặc trưng tương ứng. Và vectơ đặc trưng này sẽ là đầu vào cho mô hình đề xuất.

Luồng công việc của bước này:

1. **Trích xuất đặc trưng**
2. **Đầu vào của mô hình**
3. **Similarity calculation (tính độ tương tự)**

C. Khuyến nghị

Điểm **cosine distance** được định nghĩa như sau:

$$S_{cosine} = \frac{v_i^T v_j}{\|v_i\| \|v_j\|}$$

Điểm S_{cosine} càng lớn thì hai hình ảnh càng giống nhau.
4. Đầu ra: hình ảnh k (sản phẩm) giống với hình ảnh mục tiêu nhất.

Nội dung trình bày

Dữ liệu và đặc trưng

Để xây dựng hệ thống khuyến nghị, chúng tôi sử dụng dữ liệu hình ảnh sản phẩm của Amazon, kéo dài từ tháng 5 năm 1996 đến tháng 7 năm 2014, trong đó bao gồm 9,4 triệu sản phẩm, với tổng số 20 danh mục.

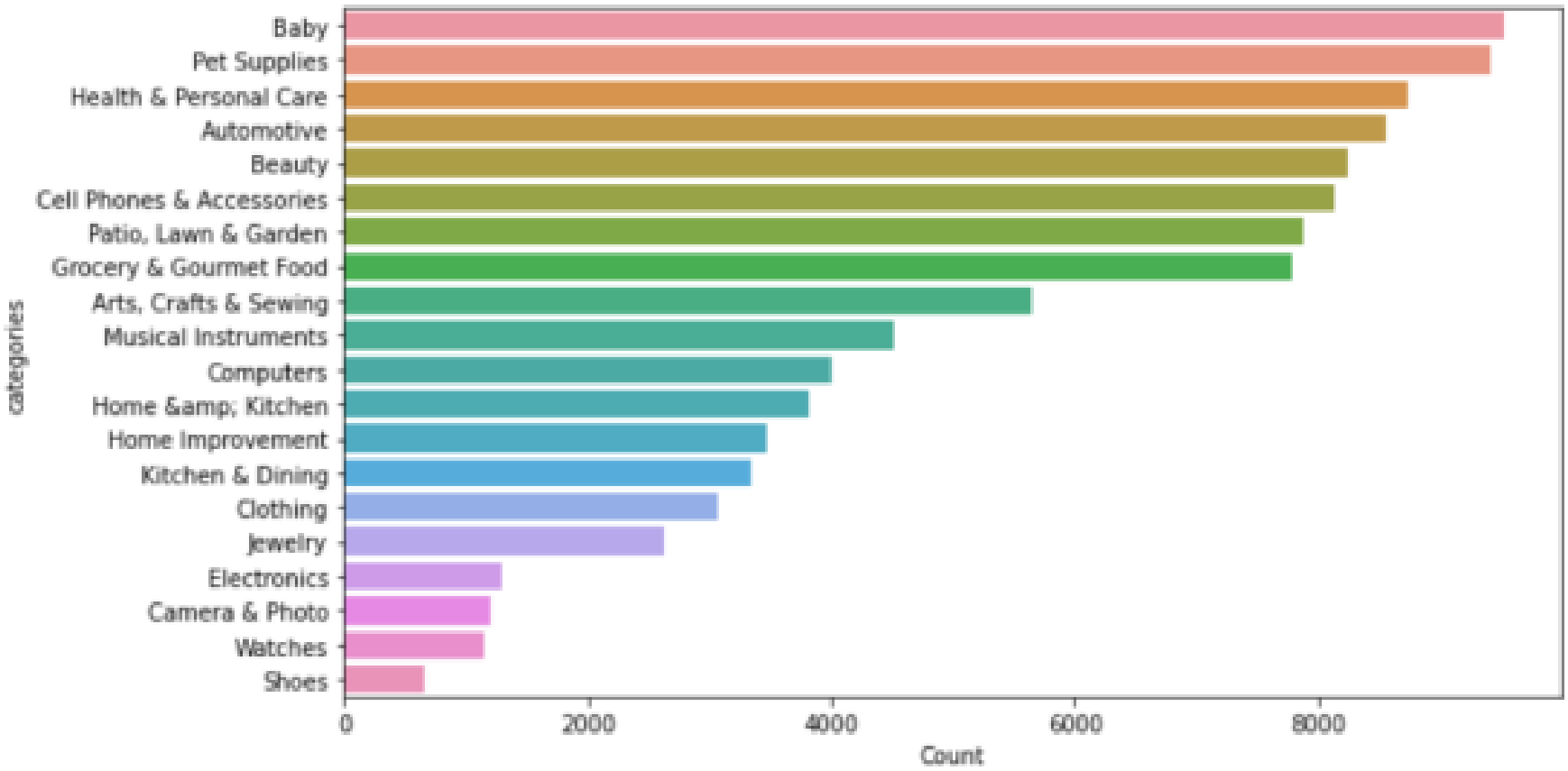
Chi tiết:

- **asin** - ID của sản phẩm, ví dụ: 0000027091
- **title** - tên của sản phẩm
- **price** - giá bằng đô la Mỹ (tại thời điểm thu thập dữ liệu)
- **imUrl** - link của hình ảnh sản phẩm
- **related** - sản phẩm liên quan (also bought, also viewed, bought together, buy after viewing)
- **salesRank** - thông tin xếp hạng bán hàng
- **brand** - tên thương hiệu
- **categories** - danh sách các danh mục sản phẩm thuộc về

Dữ liệu và đặc trưng

	categories	imUrl	price	asin	description	title
76546	Grocery & Gourmet Food	http://ecx.images-amazon.com/images/I/31L-McAf...	6.99	B0008DI8QM	Softer caramel, wonderful for centers, apple d...	Mercken's Block Vanilla Caramel
92199	Musical Instruments	http://ecx.images-amazon.com/images/I/414NKFSM...	49.95	B00006HMQ6	Take your talents to another level or embark u...	Casio LK-43 Lighted Keyboard
477	Baby	http://ecx.images-amazon.com/images/I/41Ag8SwJ...	7.48	B00005BYUL	Sassy baby Large, Medium and Small Feeding Bow...	Sassy Baby Large, Medium, And Small Feeding Bo...
52087	Cell Phones & Accessories	http://ecx.images-amazon.com/images/I/516QyZ7J...	5.77	B000TA8F80	Features: long lasting high grade cow leather,...	Cellet Motorola RAZR V3 "Posh Case" ...
111690	Shoes	http://ecx.images-amazon.com/images/I/41YCSEcm...	NaN	B0001HMC7Q	NaN	Men's LaCrosse® 18" Alphaburly Huntin...
50330	Cell Phones & Accessories	http://ecx.images-amazon.com/images/I/41OuuZNL...	11.15	B000NUTPLW	Cellet Omega Case Series provide excellent pro...	Cellet Horizontal Omega Pouch for HD2 & EV...
58770	Health & Personal Care	http://ecx.images-amazon.com/images/I/41Mte2BW...	17.56	B00028OIKI	Bluebonnet Nutrition Lycopene 20 mg - 30 Softg...	Bluebonnet Nutrition - Lycopene 20 mg. - 30 So...
59462	Health & Personal Care	http://ecx.images-amazon.com/images/I/41k78INN...	18.99	B0002DUN4I	multi+ complete, the unique food-based high-po...	Genuine Health: multi+ complete (60Tablets)
2659	Baby	http://ecx.images-amazon.com/images/I/51IL1ZSC...	NaN	B000A1AF8G	This kit includes a revised user guide, belt-r...	Britax Regent Youth Car Seat, Sahara
26977	Kitchen & Dining	http://ecx.images-amazon.com/images/I/41xXoqj9...	17.99	B005IHCGJ8	Good quality cigarette holder with integrated,...	Denicotea 20202 Ejector Lady Black and Gold Ho...

Dữ liệu và đặc trưng



Hình 7. Biểu đồ phân phối các danh mục trong tập dữ liệu

Dữ liệu và đặc trưng

Hình 8. Ví dụ về dữ liệu. Đây là ba sản phẩm từ danh mục "Cell Phones & Accessories"

Nội dung trình bày

A. Xử lý dữ liệu

Các hình ảnh thô cần được xử lý trước trước khi được sử dụng làm đầu vào của các mô hình phân loại.

Đầu tiên, một hình ảnh gốc được thay đổi kích thước thành kích thước đầu vào tiêu chuẩn của mô hình VGG, ResNet (224 × 224) hoặc mô hình AlexNet (227 × 227) (Hình 9)

Hình 9. Xử lý hình ảnh đầu vào. Bên trái là hình ảnh gốc ban đầu. Bên phải là hình ảnh đã thay đổi kích thước (224 × 224 pixel).

B. Đánh giá

Bộ dữ liệu được chia theo tỉ lệ 7:2:1 cho training, validation và test tương ứng. Sau đó, dựa vào kết quả training trên tập dữ liệu thử nghiệm và so sánh kết quả dầu ra với kết quả thực để đánh giá các mô hình.

Đánh giá bằng cách tính toán độ chính xác của phân loại:

$$Accuracy = \frac{correctly-classified-images}{images-in-validation-dataset}$$

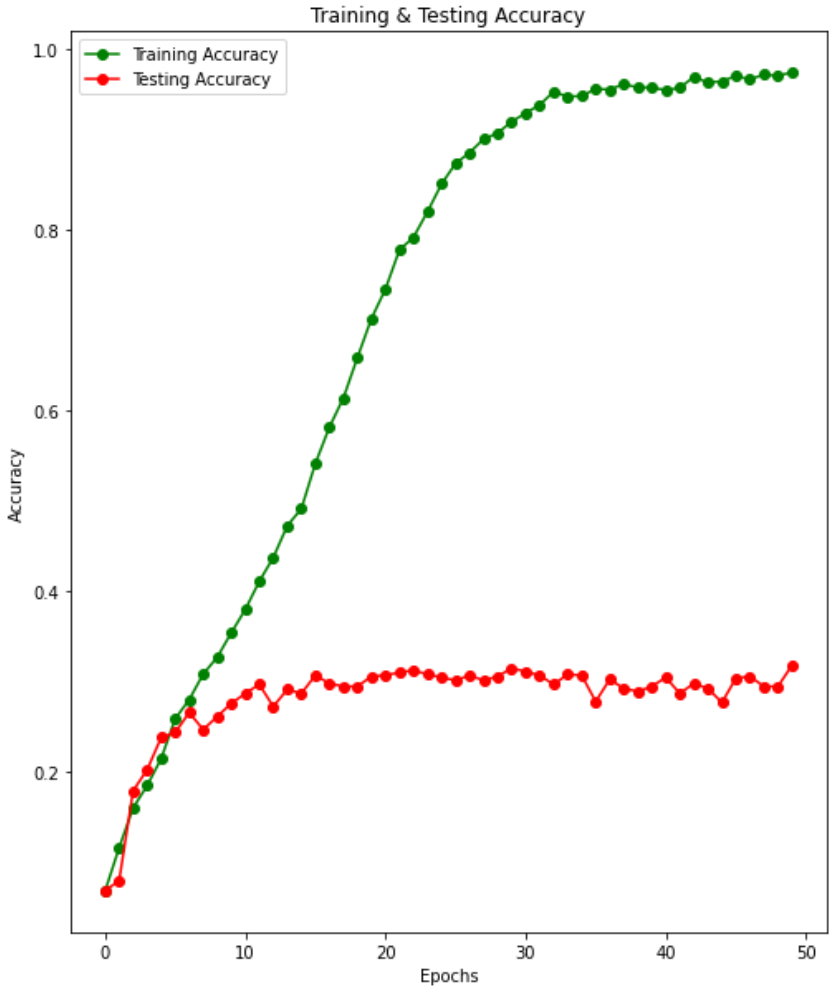
C. Phân loại

Đối với nhiệm vụ phân loại, phân loại các loại hình ảnh sản phẩm so với mô hình phân loại tuyến tính theo mô hình cơ sở (mô hình **SVM**) thông qua việc đào tạo các Mạng Nơ-Ron tích chập (**AlexNet**, **VGG16**, **VGG19** và **ResNet50**).

Model	**Train acc.**	**Valid acc.**	**Test acc.**
SVM (baseline)	0.29	0.25	0.22
AlexNet	0.7927	0.2948	0.3020
VGG16	0.9768	0.3861	0.3717
VGG19	0.9986	0.4472	0.4300
VGG19 have BatchNormalization	0.9994	0.4549	0.4400
ResNet-50	0.9989	0.4872	0.4820

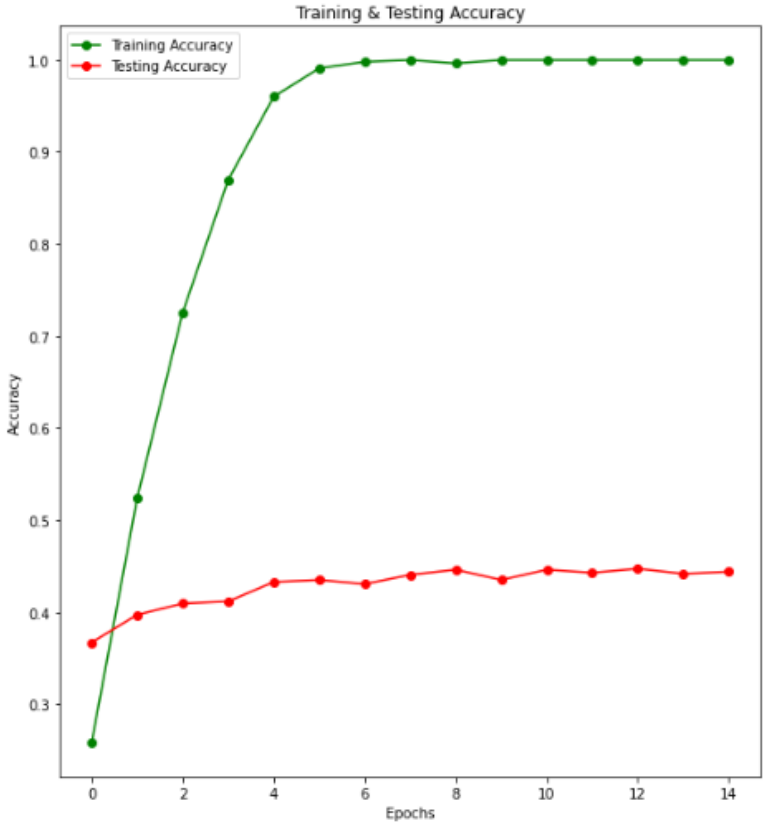
Bảng 1. Kết quả độ chính xác của các mô hình

C. Phân loại



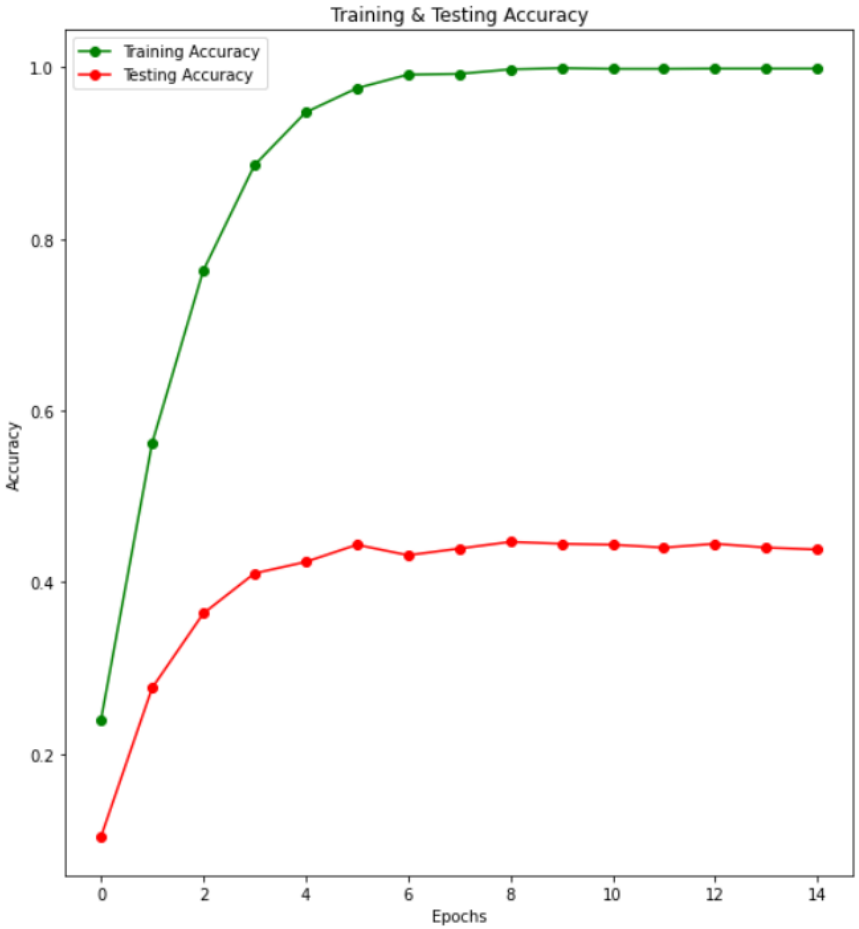
VGG 16

C. Phân loại



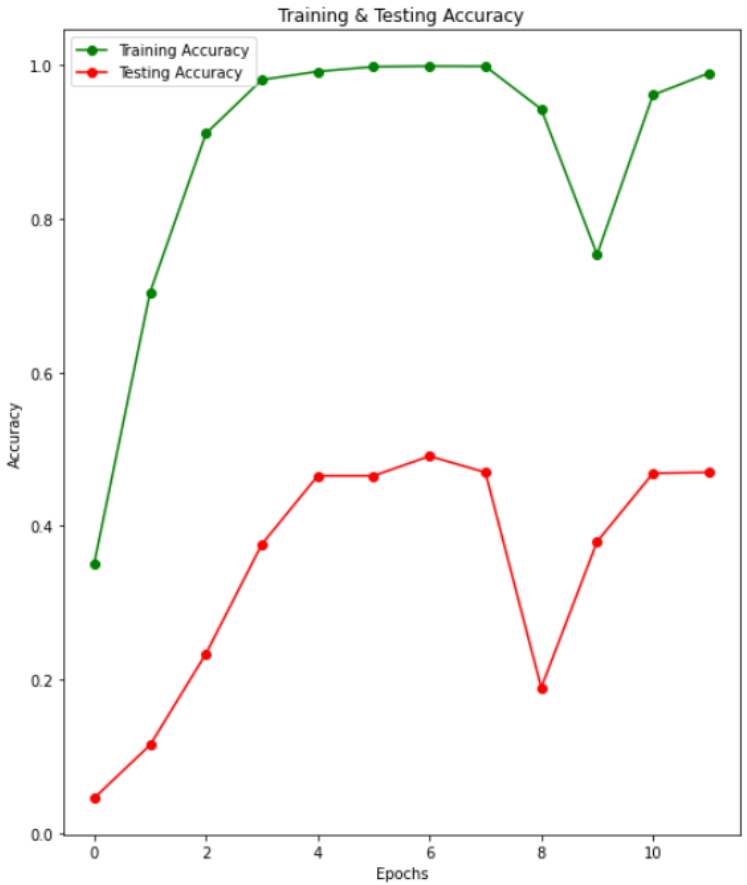
VGG 19

C. Phân loại



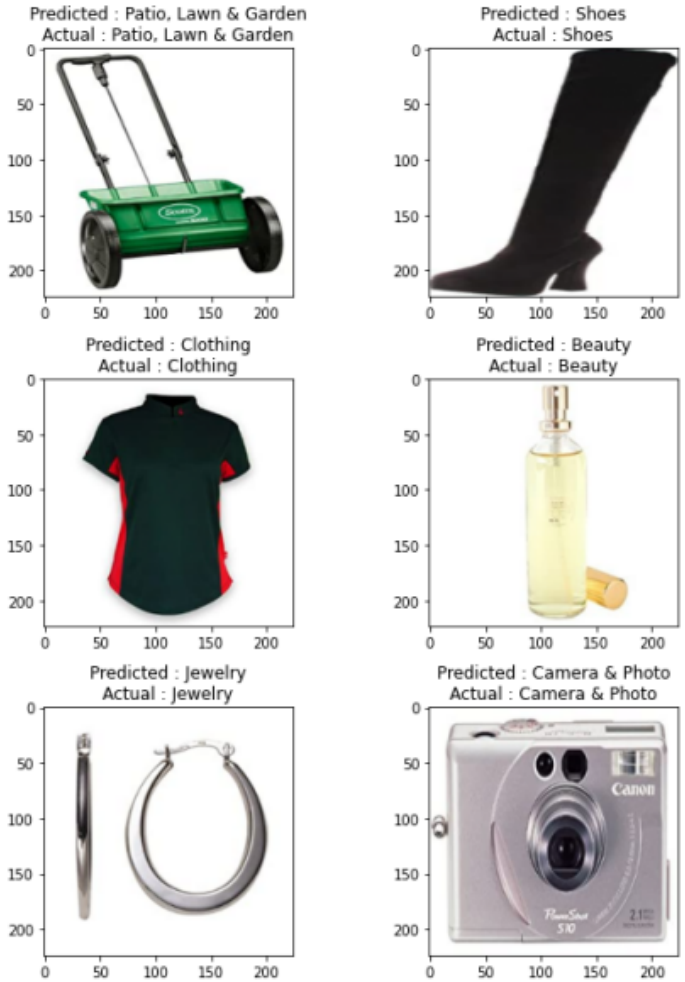
VGG 19 have BatchNormalization

C. Phân loại



ResNet 50

C. Phân loại



Hình 11. Hình ảnh các sản phẩm được phân loại đúng

C. Phân loại

Hình 12. Hình ảnh các sản phẩm được phân loại không đúng

D. Khuyến nghị

Hình 13. Ví dụ về kết quả hệ thống đề xuất của chúng tôi

D. Khuyến nghị

Hình 14. Ví dụ về kết quả hệ thống đề xuất của chúng tôi

Nội dung trình bày

Kết luận

Problem: Hiện tại chúng tôi chỉ sử dụng 20 danh mục khi thực hiện phân loại. Tuy nhiên, các sản phẩm trong danh mục khác nhau rất nhiều, điều này giải thích cho việc phân loại của chúng tôi có độ chính xác thấp.
Bên cạnh đó, chúng tôi cũng muốn thử các mạng thần kinh sâu hơn như DenseNet.

thanks for watching