

FASHION IMAGE SEARCH ENGINE

Môn học: Xử lý ảnh và ứng dụng

Lớp: CS406.P11

GVHD: Cáp Phạm Đình Thăng

Nhóm thực hiện:

Nguyễn Công Nguyên – 21521200
Trần Lê Bảo Trung - 21521598

Nội dung

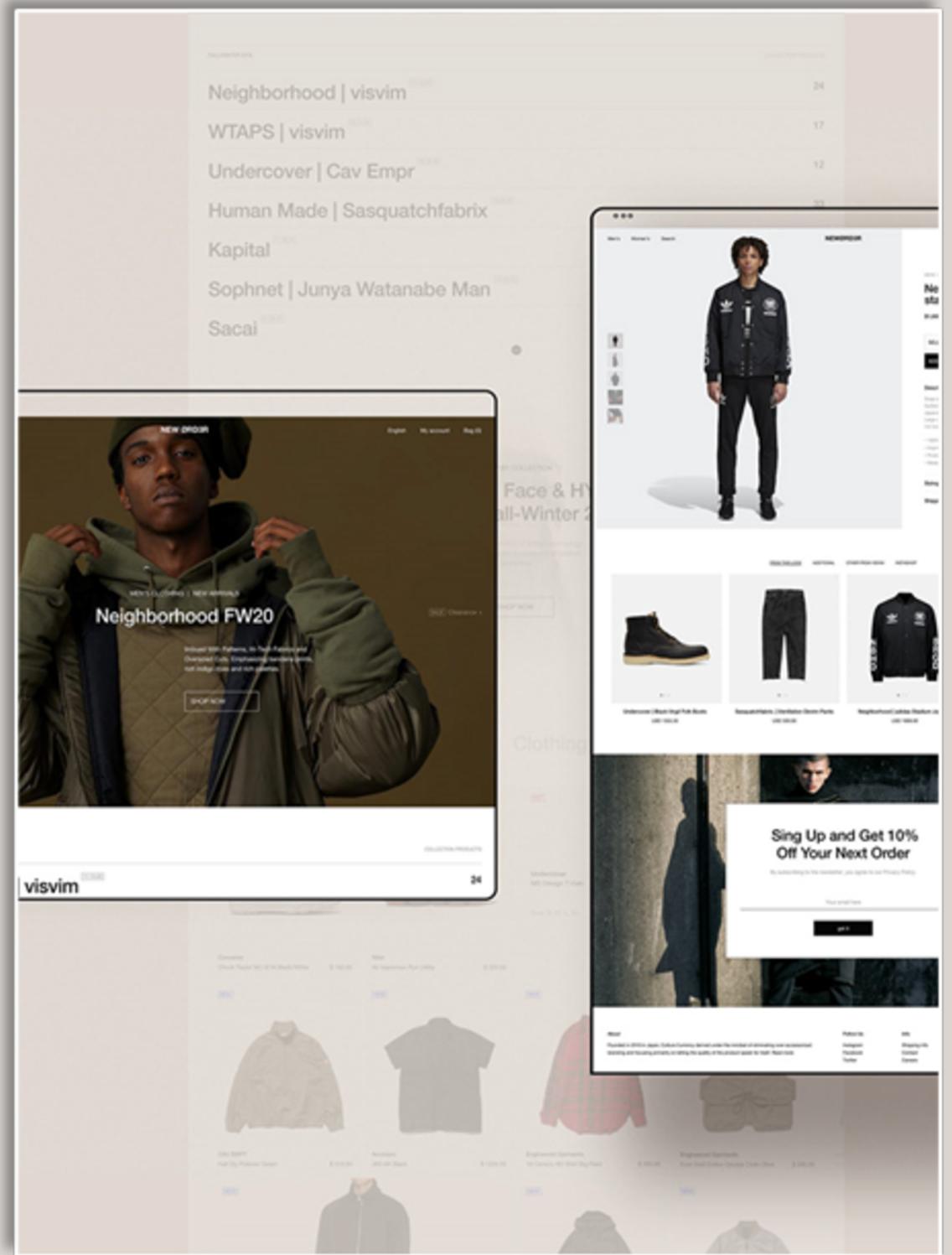
- 01 Giới thiệu đề tài
- 02 Phát biểu bài toán
- 03 Phương pháp
- 04 Thực nghiệm
- 05 Tổng kết
- 06 Demo

1. Giới thiệu đề tài

Tổng quan đề tài

1. Bối cảnh

- Việc mua sắm trực tuyến ngày càng phổ biến, đặc biệt trong lĩnh vực thời trang.
- Người tiêu dùng muốn định vị những mẫu sản phẩm tương tự hình ảnh họ nhìn thấy trên Internet.
=> Một hệ thống truy vấn hình ảnh liên quan trong lĩnh vực thời trang sẽ hỗ trợ cho người tiêu dùng trong trường hợp đó.



Tổng quan đề tài

2. Mục đích và ý nghĩa



Tăng cường trải nghiệm người dùng

Hệ thống truy vấn hình ảnh giúp người dùng dễ dàng tìm kiếm sản phẩm thời trang trực tuyến chỉ bằng cách cung cấp hình ảnh, tiết kiệm thời gian và công sức.



Nâng cao hiệu quả kinh doanh

Cải thiện khả năng gợi ý sản phẩm, tăng tỷ lệ chuyển đổi và doanh thu cho các cửa hàng thời trang trực tuyến.



Xu hướng cá nhân hóa

Đáp ứng nhu cầu cá nhân của người dùng thông qua gợi ý sản phẩm dựa trên sở thích và phong cách cá nhân.

2. Phát biểu bài toán

- Bài toán của nhóm là bài toán truy vấn hình ảnh (với các hình ảnh trong lĩnh vực thời trang).
- Người dùng chỉ cần cung cấp một hình ảnh của sản phẩm mong muốn, chẳng hạn như giày, quần áo, hoặc ví. Kết quả trả về sẽ là một danh sách các sản phẩm tương tự hoặc phù hợp, danh sách sẽ được sắp xếp theo thứ tự giảm dần về độ tương đồng.



INPUT

Một hình ảnh truy vấn chứa sản phẩm thời trang mà người dùng quan tâm (quần áo, giày dép, ví,...).



OUTPUT

Một danh sách được xếp hạng các hình ảnh có độ tương đồng cao nhất với hình ảnh truy vấn, với hình ảnh giống nhất đứng đầu danh sách.

>>

INPUT

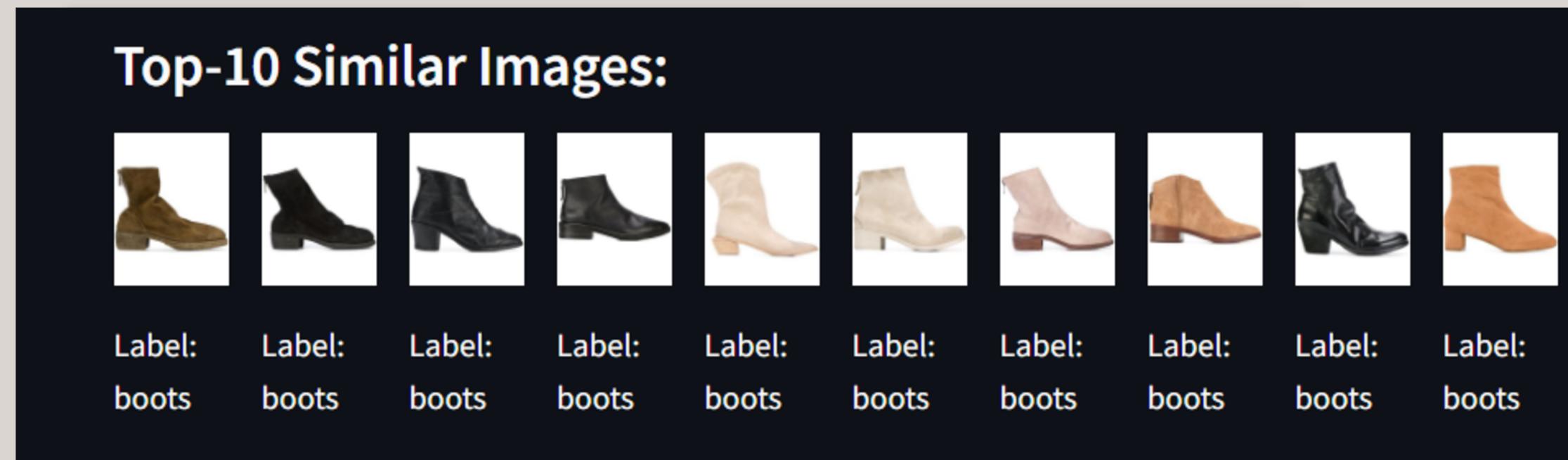
Một hình ảnh truy vấn chứa sản phẩm thời trang mà người dùng quan tâm (quần áo, giày dép, ví,...).



>>

OUTPUT

Một danh sách được xếp hạng các hình ảnh có độ tương đồng cao nhất với hình ảnh truy vấn, với hình ảnh giống nhất đứng đầu danh sách.



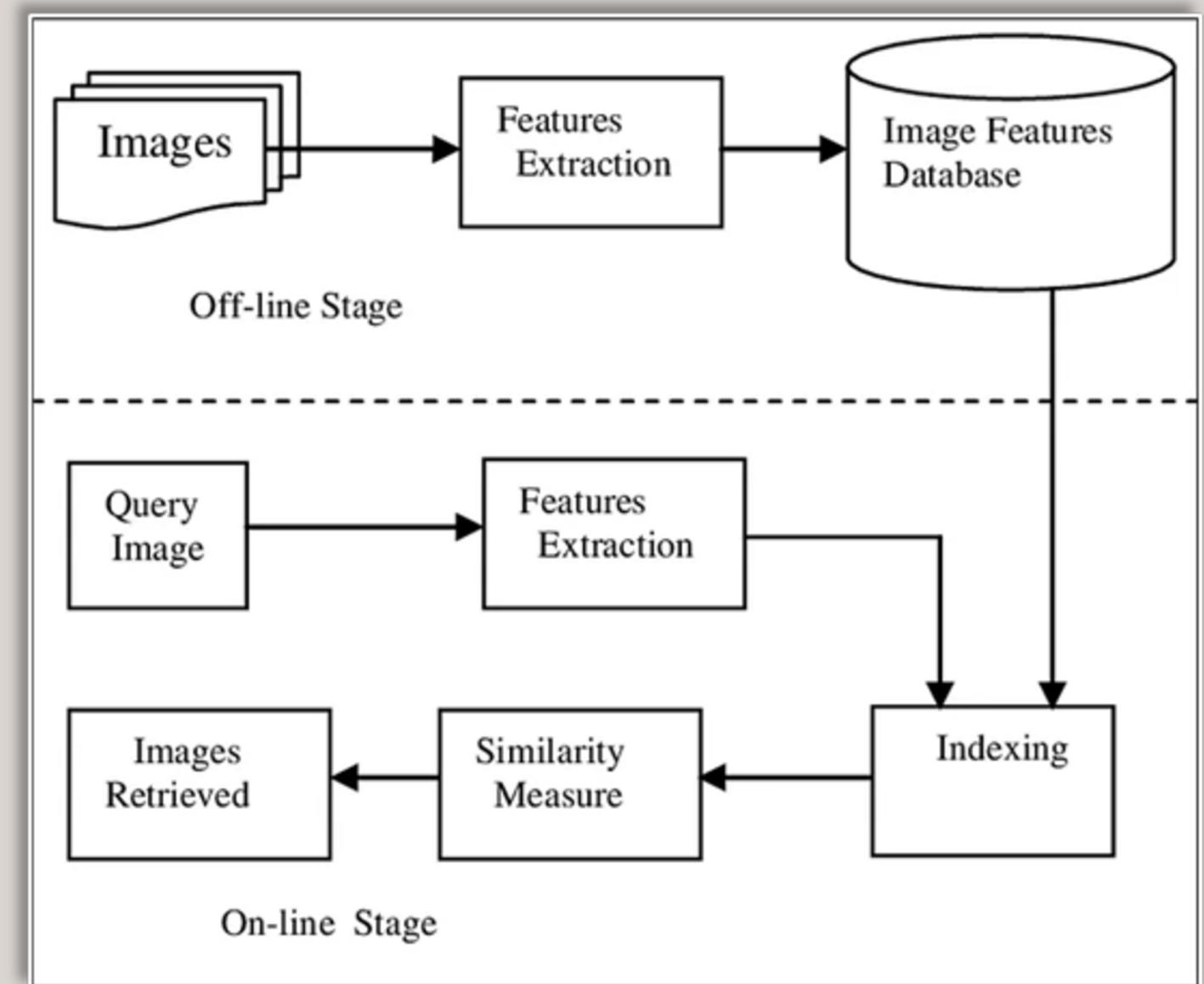
Hình 2. Minh họa input bài toán

Hình 3. Minh họa output tương ứng

3. Phương pháp

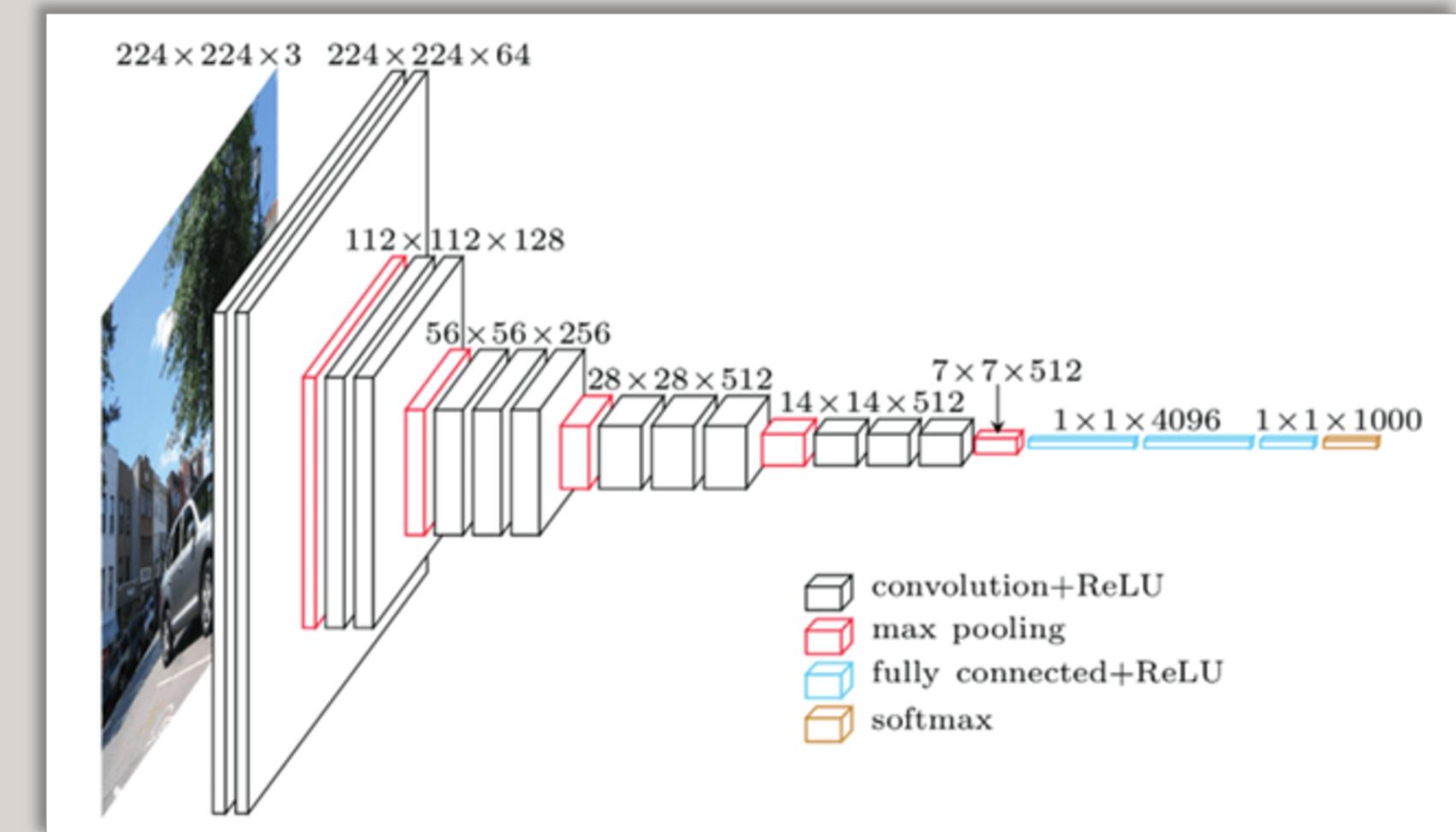
3.1 Hệ thống truy vấn hình ảnh

- Mục tiêu chính: Tính toán độ tương đồng giữa hình ảnh truy vấn và cơ sở dữ liệu để trích xuất hình ảnh tương tự nhất.
- Các hình ảnh trong cơ sở dữ liệu sẽ được trích xuất đặc trưng và lưu trữ lại.
- Khi có yêu cầu truy vấn, độ tương đồng giữa hình ảnh truy vấn và hình ảnh trong cơ sở dữ liệu được tính toán dựa trên các đặc trưng.



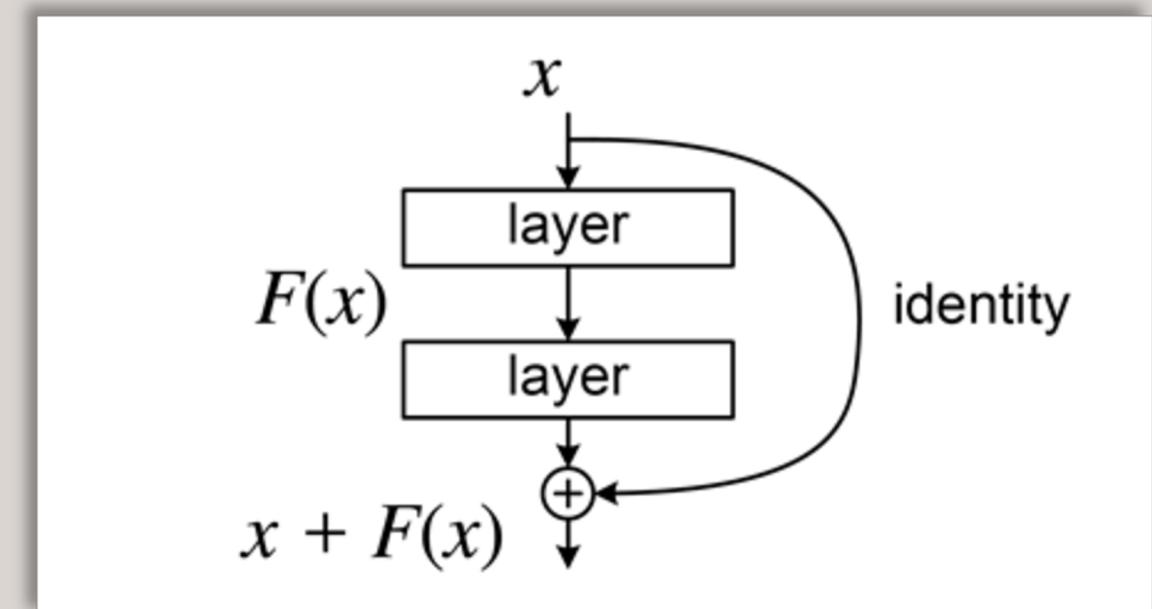
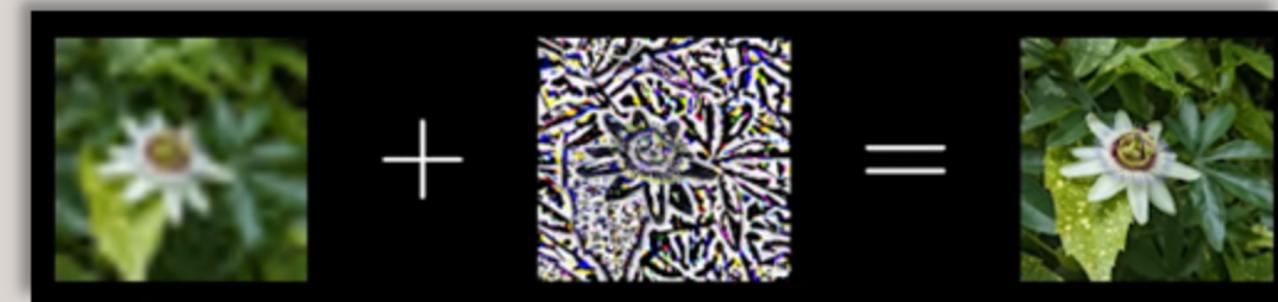
3.2 VGG16

- VGG16 có 13 lớp Convolutional để trích xuất đặc trưng hình ảnh, 5 lớp Max Pooling giảm kích thước đầu ra, và 3 lớp fully connected ở cuối để tổng hợp các đặc trưng.
- Output của fully-connected layer được đưa vào lớp Softmax để phân loại.
- Đặc trưng được trích xuất ở lớp fully connected trước lớp phân loại.



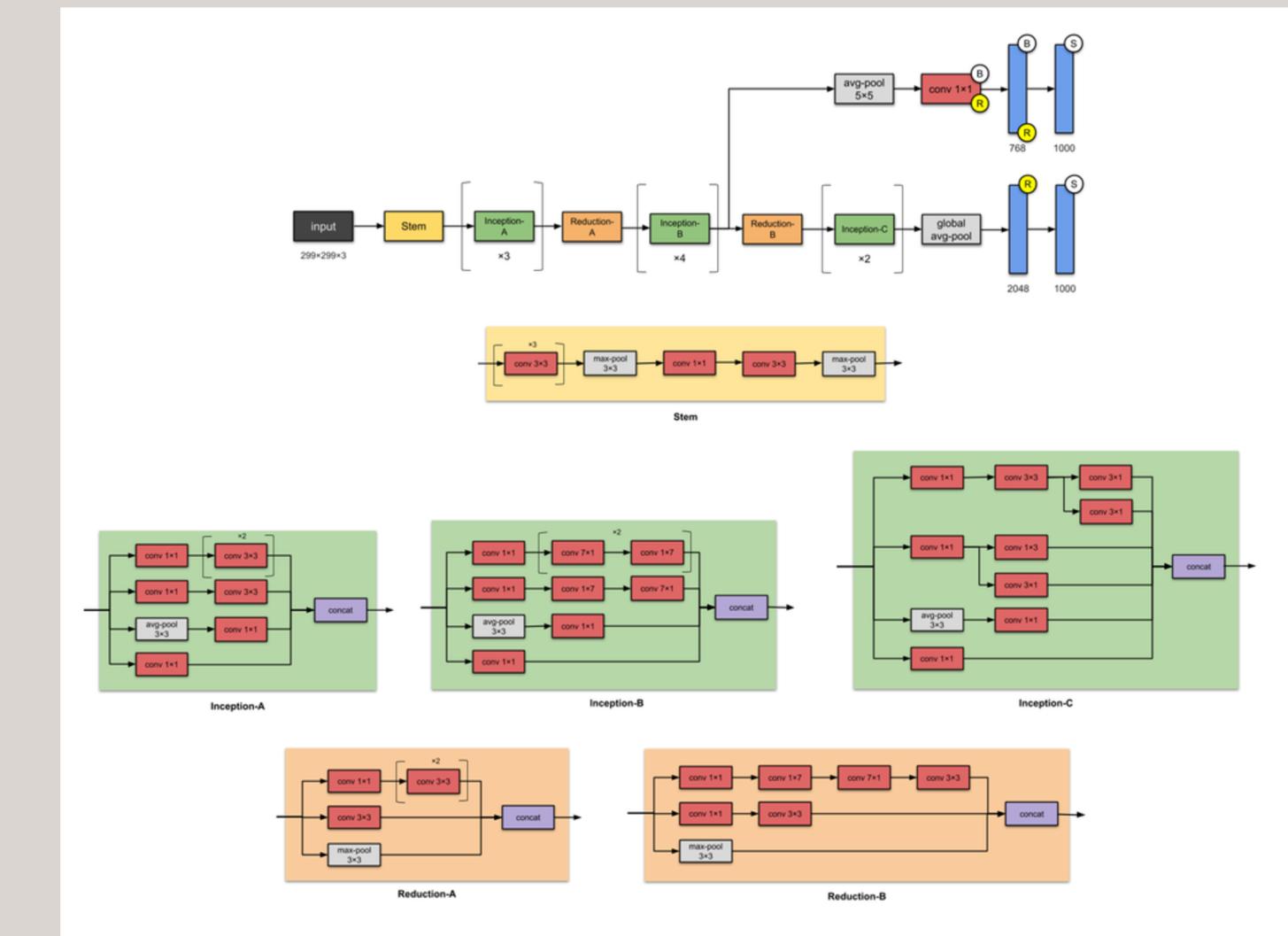
3.3 Resnet 50

- Ý tưởng: ResNet học cách giữ lại thông tin đầu vào và chỉ học thêm các thay đổi nhỏ qua từng lớp để đưa nó đến gần hơn với giá trị đầu ra thực sự.
- Residual Blocks
 - ResNet nhóm một số lớp tích chập lại và gọi nó là một Residual Block.
 - Dữ liệu đầu vào của một Block không chỉ đi qua các lớp bên trong để học những thay đổi mà còn được truyền trực tiếp (identity mapping) đến đầu ra của Block.
- Shortcut Connections: Các kết nối từ đầu vào đến đầu ra của một Block, duy trì tín hiệu và giảm thiểu mất mát thông tin.
- Đặc trưng sẽ được trích xuất ở lớp Average Pooling.



3.3 Inception V3

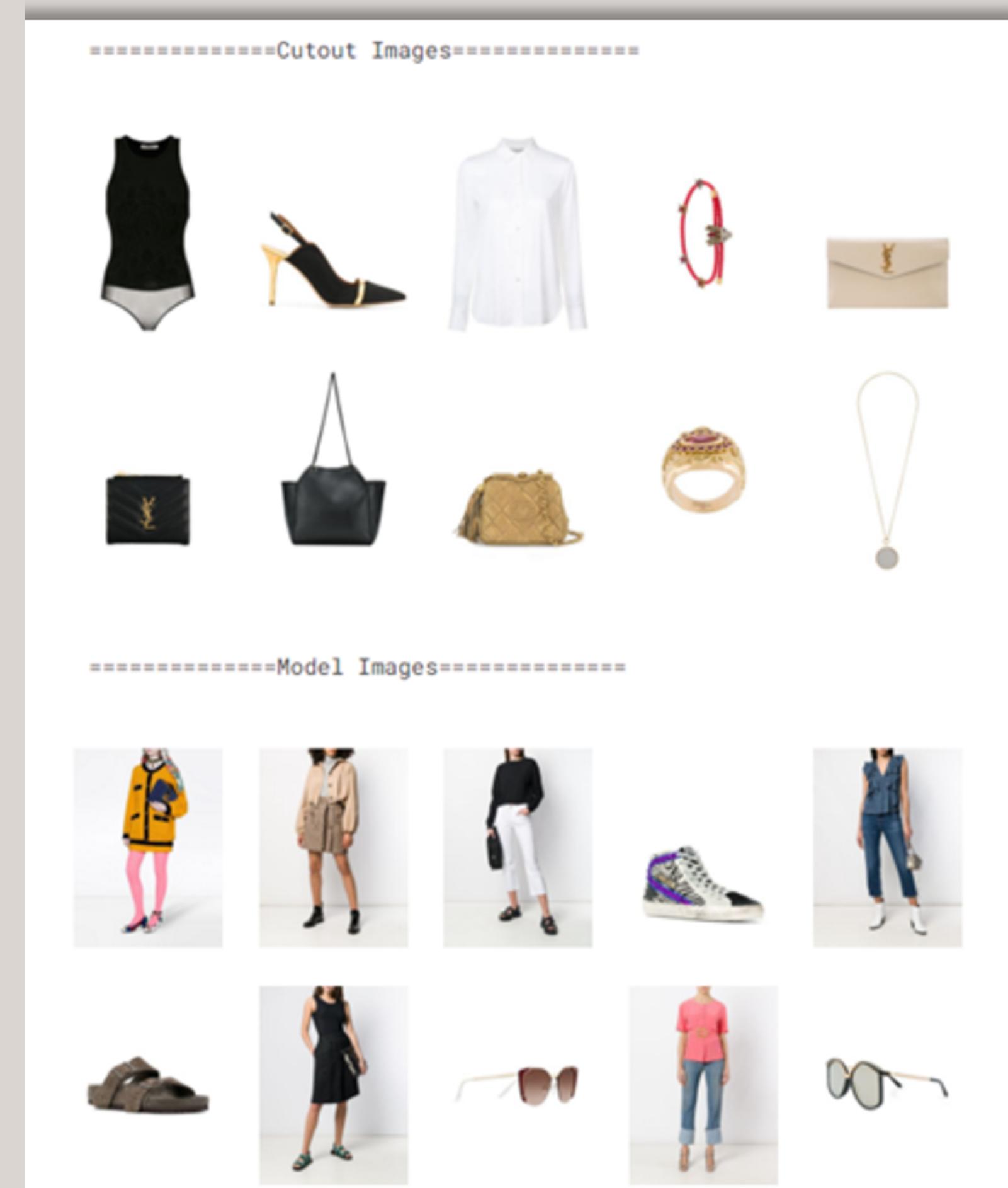
- Inception-V3 là kế thừa của Inception-V1, bao gồm 24 triệu tham số. Toàn bộ các layer tích chập của Inception-V3 được theo sau bởi một layer batch normalization và một ReLU activation.
- Batch normalization là kỹ thuật chuẩn hóa đầu vào theo từng minibatch tại mỗi layer theo phân phối chuẩn hóa $N(0,1)$, giúp cho quá trình huấn luyện thuật toán nhanh hơn.
- Trong Inception-V3, đặc trưng được trích xuất tại lớp Average Pooling.



4. Thực nghiệm

4.1 Bộ dữ liệu

- Bộ dữ liệu Farfetch Listings chứa 187,620 hình ảnh thời trang từ trang web FarFetch, tất cả các ảnh đều lưu định dạng jpg.
- Trong đó có 2 thư mục cutout-img và model-img, với thư mục cutout-img chứa hình thời trang đã được xóa phông, thu mục model-img chứa hình ảnh người mẫu.
- Bộ dữ liệu sau khi đã qua xử lý tạo thành 37 nhãn đại diện cho các danh mục thời trang khác nhau.
- Dữ liệu được chia theo tỉ lệ train/validation/test là 8/1/1.



4.2 Độ đo

- mAP đánh giá hiệu suất của hệ thống trong việc truy xuất hình ảnh có liên quan từ cơ sở dữ liệu, dựa trên một truy vấn của người dùng.
- Công thức tính độ chính xác trung bình (Average Precision):

$$AP = \frac{1}{\text{total relevant documents}} \sum_{k=1}^n (\text{Precision at } k * \text{rel}(k))$$

- $\text{rel}(k)$ là một hàm chỉ số bằng 1 nếu mục ở hạng k là một tài liệu có liên quan, bằng 0 nếu không phải. Lưu ý rằng trung bình được tính trên các tài liệu có liên quan trong top-k tài liệu được trả về, và các tài liệu có liên quan mà không được trả về sẽ có điểm độ chính xác bằng 0

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP_q$$

- Q là số lượng truy vấn. Điểm mAP càng cao, hiệu suất của hệ thống CBIR càng tốt. mAP là một độ đo phổ biến được sử dụng để đánh giá hệ thống CBIR, vì nó xem xét tính liên quan của mỗi hình ảnh được truy xuất và cung cấp một đo lường toàn diện về hiệu suất của hệ thống qua nhiều truy vấn.

4.3 Kết quả thực nghiệm

Top-k	InceptionV3	ResNet50	VGG16
Top 1	0.7938	0.8122	0.7885
Top 5	0.8355	0.8550	0.8329
Top 10	0.8241	0.8433	0.8251

Bảng 1: Kết quả thực nghiệm mAP của các mô hình trên bộ dữ liệu Farfetch-Listing

Từ bảng trên, có thể thấy rằng ResNet50 đạt điểm mAP cao nhất ở các top-k so với InceptionV3 và VGG16. Cụ thể:

- Ở top 1, ResNet50 đạt độ chính xác 0.8122, cao hơn một chút so với InceptionV3 (0.7938) và VGG16 (0.7885).
- Ở top 5, ResNet50 cũng đạt điểm cao nhất với mAP là 0.8550, vượt trội hơn so với InceptionV3 (0.8355) và VGG16 (0.8329).
- Ở top 10, ResNet50 vẫn giữ vị trí dẫn đầu với mAP là 0.8433, trong khi InceptionV3 (0.8241) và VGG16 (0.8251) gần như ngang bằng nhau.

5. Tổng kết

5.1 Kết luận

- Hiệu quả của mô hình CNN: ResNet50, VGG16, và InceptionV3 đạt hiệu suất cao trong bài toán truy vấn hình ảnh thời trang. ResNet50 vượt trội với mAP cao nhất nhờ kiến trúc residual connections.
- Khoảng cách Euclidean: Là phương pháp đơn giản, hiệu quả để đo độ tương đồng trong không gian đặc trưng.
- Indexing: Dù chưa hoàn thiện, đã đặt nền tảng cho việc lưu trữ đặc trưng, tối ưu hóa truy vấn và cải thiện hiệu suất.

5.2 Hướng phát triển

- Tối ưu hóa chỉ mục: Sử dụng Faiss, Annoy, hoặc Elasticsearch để tăng tốc truy vấn và mở rộng hệ thống.
- Thử nghiệm mô hình mới: EfficientNet, MobileNet để nâng cao hiệu suất và trải nghiệm người dùng.
- Tích hợp tính năng tương tác: Tìm kiếm theo màu sắc, xu hướng thời trang, và gợi ý sản phẩm.
- So sánh phương pháp đo lường: Nghiên cứu thêm Cosine Similarity, Jaccard Similarity để cải thiện độ chính xác.

6. Demo

**Thank
You**