
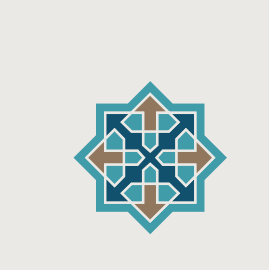


Image Retrieval



Introduction

Image retrieval là một lĩnh vực khá nổi bật trong thị giác máy tính, liên quan đến việc **tìm kiếm và truy xuất** các hình ảnh **dựa trên các đặc trưng** của chúng. Nó đang trở thành một chủ đề quan tâm đến trong thị giác máy tính vì sự phát triển của công nghệ số.



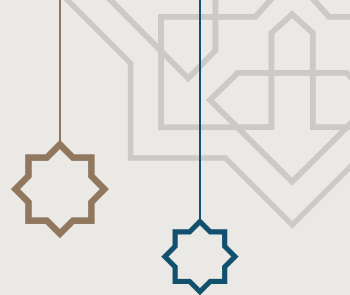
Động lực nghiên cứu

Động lực khoa học

Là một lĩnh vực nghiên cứu quan trọng trong khoa học máy tính và trí tuệ nhân tạo.

Liên quan đến việc xử lý dữ liệu ảnh và tìm kiếm thông tin trong ảnh, đó là những vấn đề rất thú vị và phức tạp.

Các nhà nghiên cứu trong lĩnh vực này đang phát triển các phương pháp mới để cải thiện hiệu quả và độ chính xác của hệ thống tìm kiếm ảnh. Họ cũng đang tìm hiểu cách để ứng dụng các công nghệ mới nhất như học sâu (deep learning) vào nghiên cứu của họ.

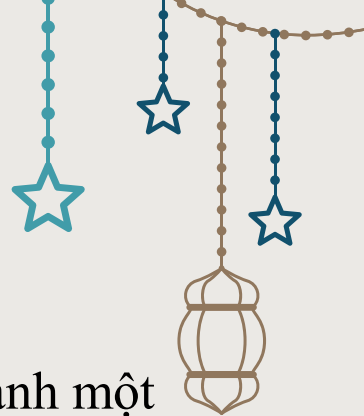


Động lực nghiên cứu

Về mặt thực tiễn

Với sự phát triển của công nghệ và internet, hình ảnh đã trở thành một phương tiện truyền thông phổ biến và quan trọng trong cuộc sống hàng ngày. Việc phát triển các công nghệ tìm kiếm ảnh sẽ giúp chúng ta tìm kiếm các hình ảnh cần thiết nhanh chóng và chính xác hơn.

Điều này rất hữu ích cho các ứng dụng thực tế như: tìm kiếm sản phẩm trên các trang web thương mại điện tử, tìm kiếm thông tin y học từ các hình ảnh y khoa, tìm kiếm thông tin về môi trường từ các hình ảnh địa lý v.v.



Challenge

Hình ảnh có thể đến từ nhiều nguồn khác nhau và có thể được chụp từ nhiều góc độ khác nhau. Hình ảnh cũng có thể **bị nhiễu** và **mất chất lượng**, gây ảnh hưởng đến kết quả truy xuất.

Vấn đề xử lý ảnh số lượng **lớn** và **đa dạng**.

Các mạng học sâu hiện có được đào tạo cho việc phân loại hình ảnh nên việc trích xuất đặc trưng trở thành một thách thức trong bài toán truy vấn.



Challenge

Các phương pháp tìm kiếm phổ biến hiện nay gồm: tìm theo từ khóa (Text-based Image Retrieval), tìm theo nội dung (Content-based Image Retrieval) và tìm theo ngữ nghĩa (Semantic-based Image Retrieval)



Phát biểu bài toán

a) Input:

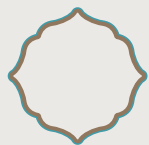
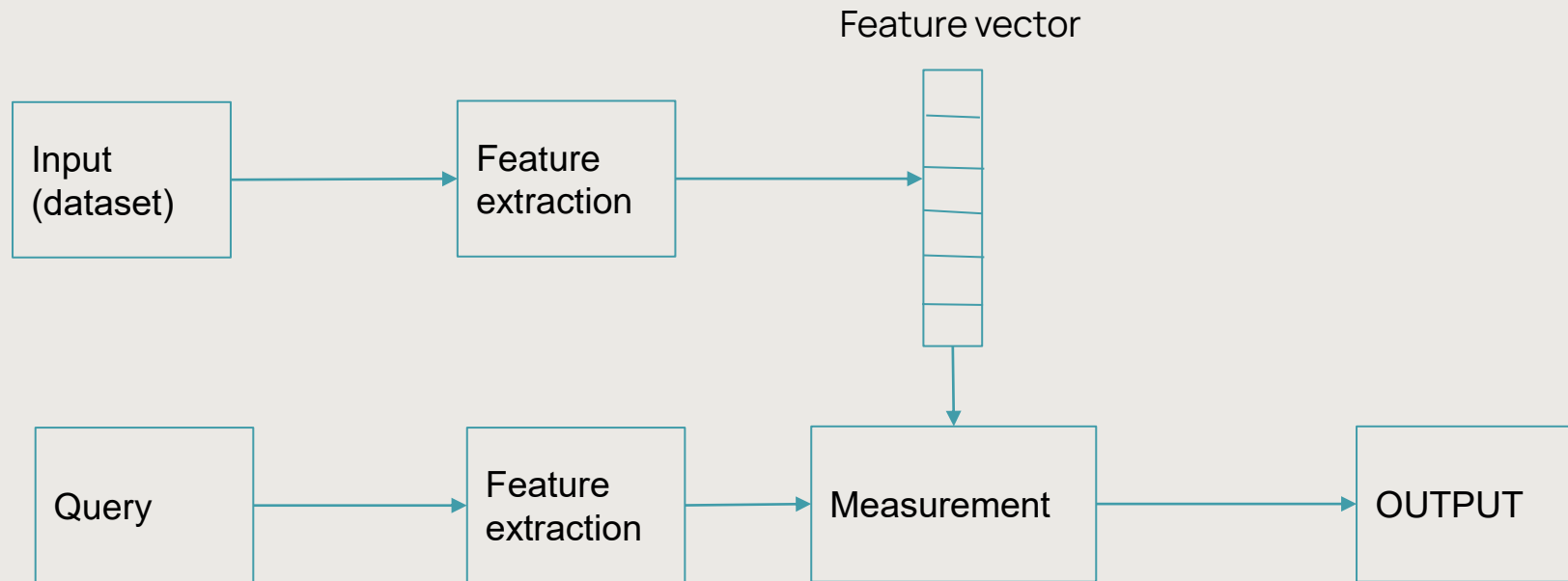
Dữ liệu đầu vào của bài toán Image Retrieval thường có 3 dạng: query image, văn bản mô tả hoặc là cả query image và văn bản mô tả.

b) Output:

Tập hợp các hình ảnh từ cơ sở dữ liệu tương đồng với query image nhất hoặc gần với mô tả văn bản nhất.



Phân tích các bài toán thành phần



Content - based image retrieval

Input: Ảnh đầu vào.

Output: Các ảnh tương tự với ảnh đầu vào theo độ tương đồng về nội dung, ví dụ như màu sắc, hình dạng, kích thước, texture, v.v.



Giải quyết bài toán

1. Bag of visual words

Phương pháp **truyền thống** trong bài toán CBIR

Đây là một phương pháp dựa trên **đặc trưng cục bộ** và áp dụng kỹ thuật **clustering** để biểu diễn hình ảnh.

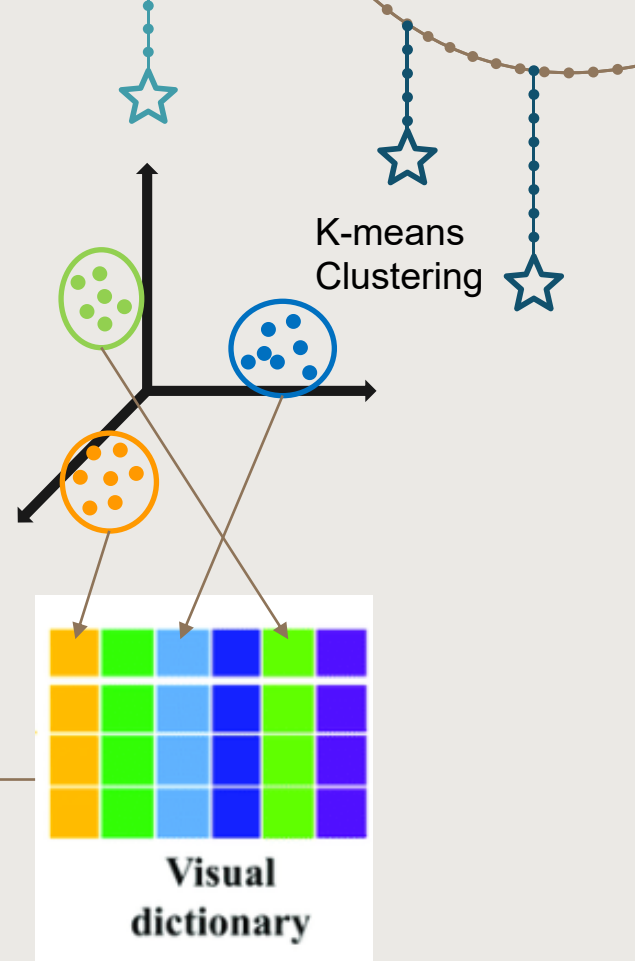




Feature extract



Clustering

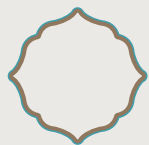


Query image

Feature extraction

Measurement with K represent

Output: image list



Giải quyết bài toán

2. Sử dụng deep learning

Framework deep learning được đề xuất cho CBIR, bao gồm hai giai đoạn:

- (i) huấn luyện một mô hình học sâu từ một bộ sưu tập lớn dữ liệu huấn luyện.
- (ii) áp dụng mô hình học sâu đã được huấn luyện để học các biểu diễn đặc trưng cho các tác vụ CBIR trong một miền dữ liệu mới.

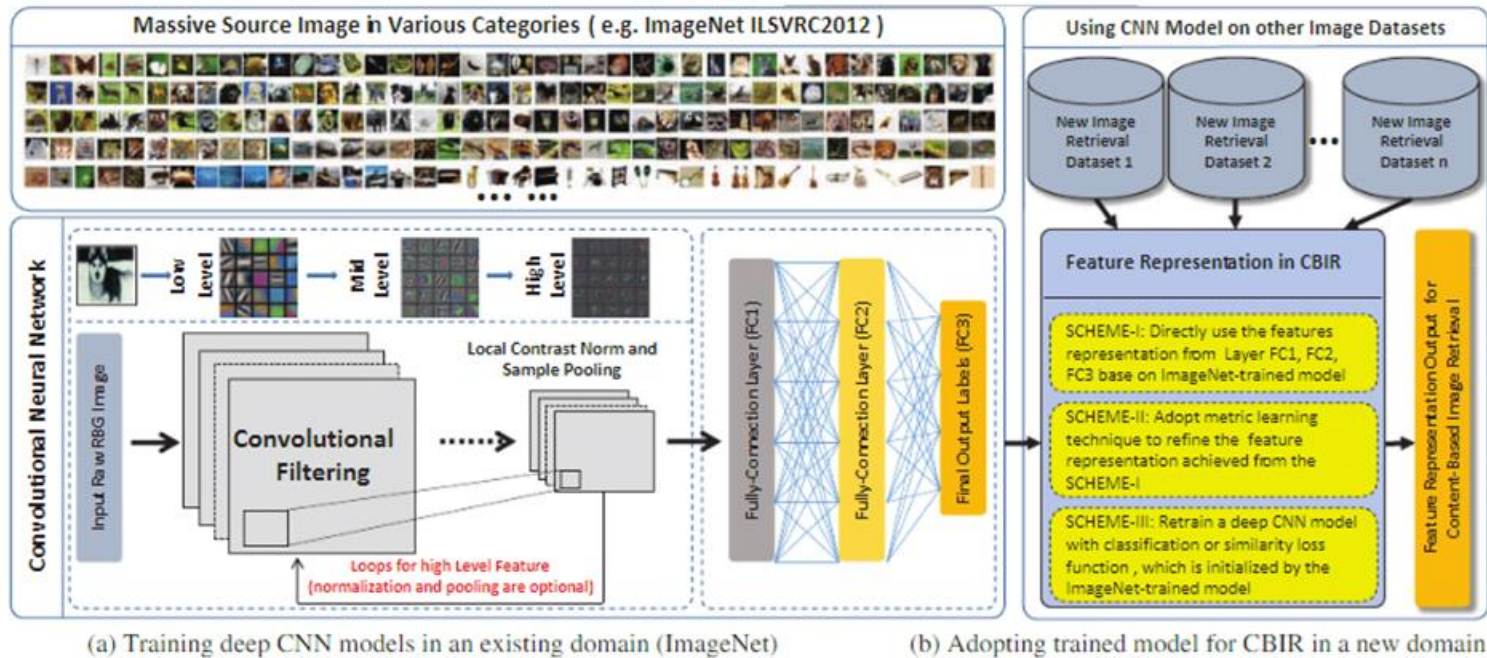
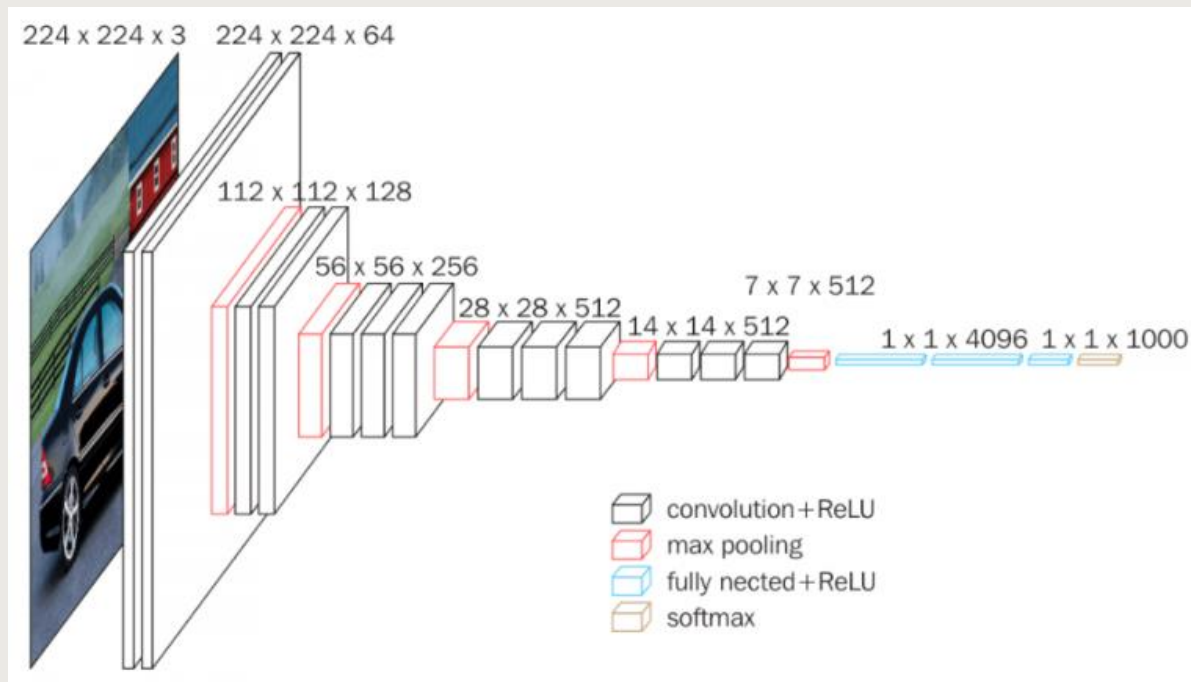


Figure 1: A Framework of Deep Learning with Application to Content-based Image Retrieval.

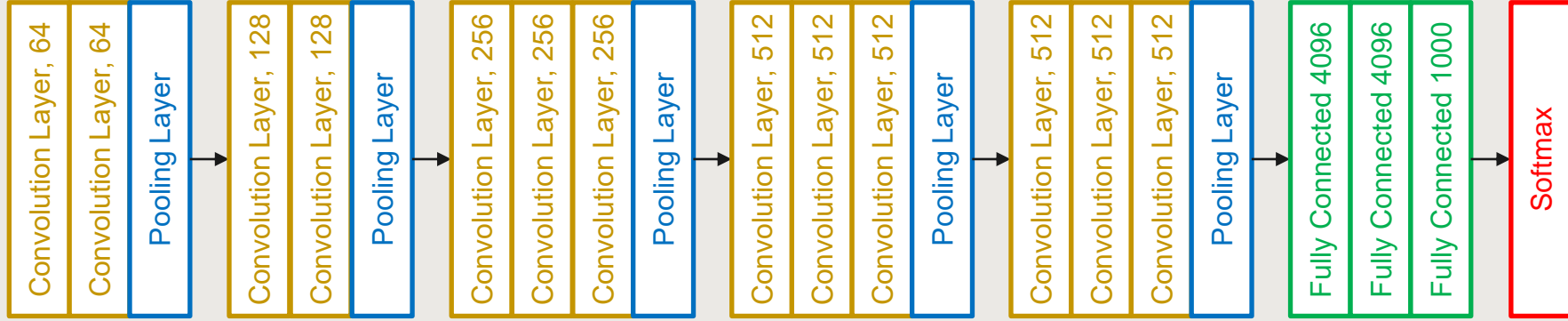
Feature extraction

VGG - 16



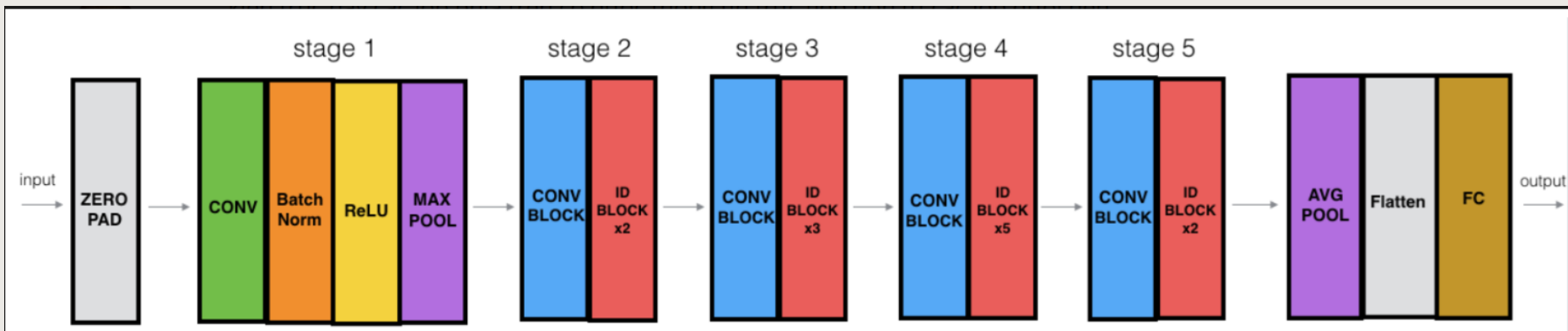
Feature extraction

VGG - 16



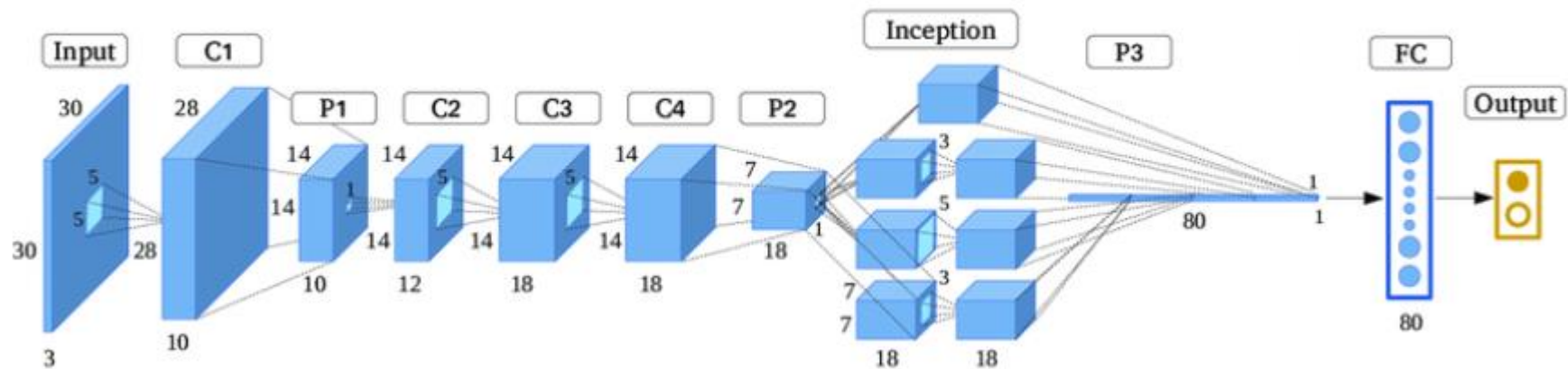
Feature extraction

ResNet

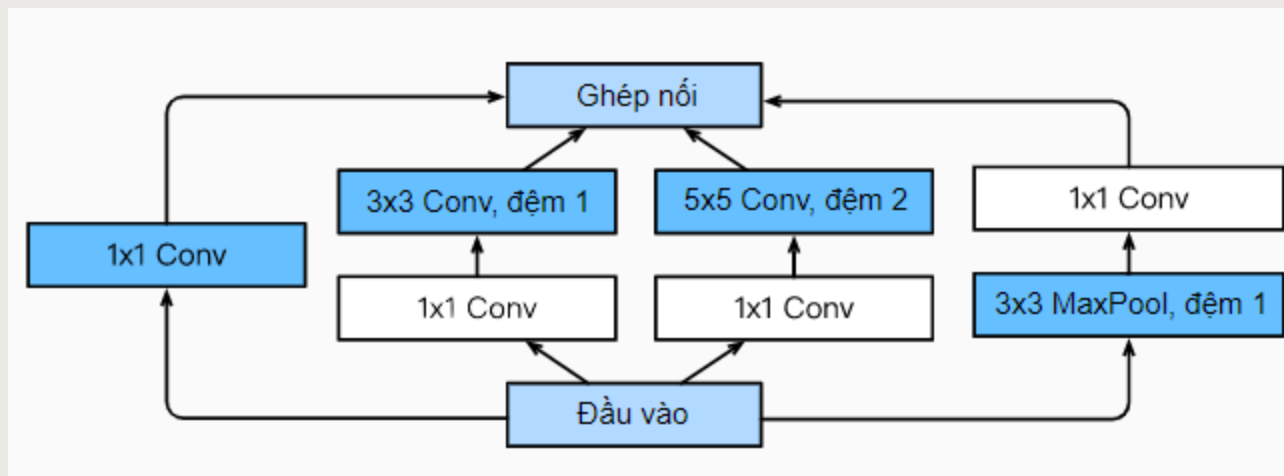


Feature extraction

Googlenet - Inception v1

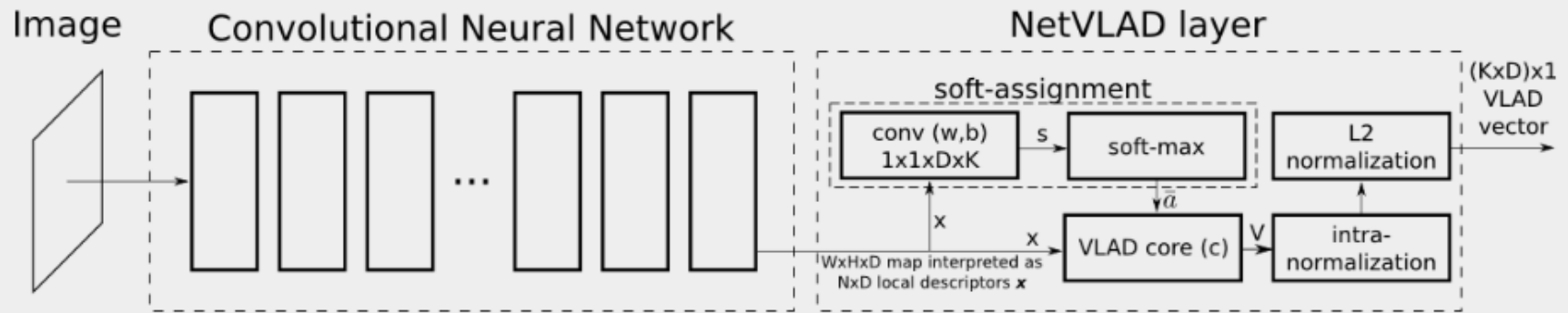


Inception Structure in Googlenet





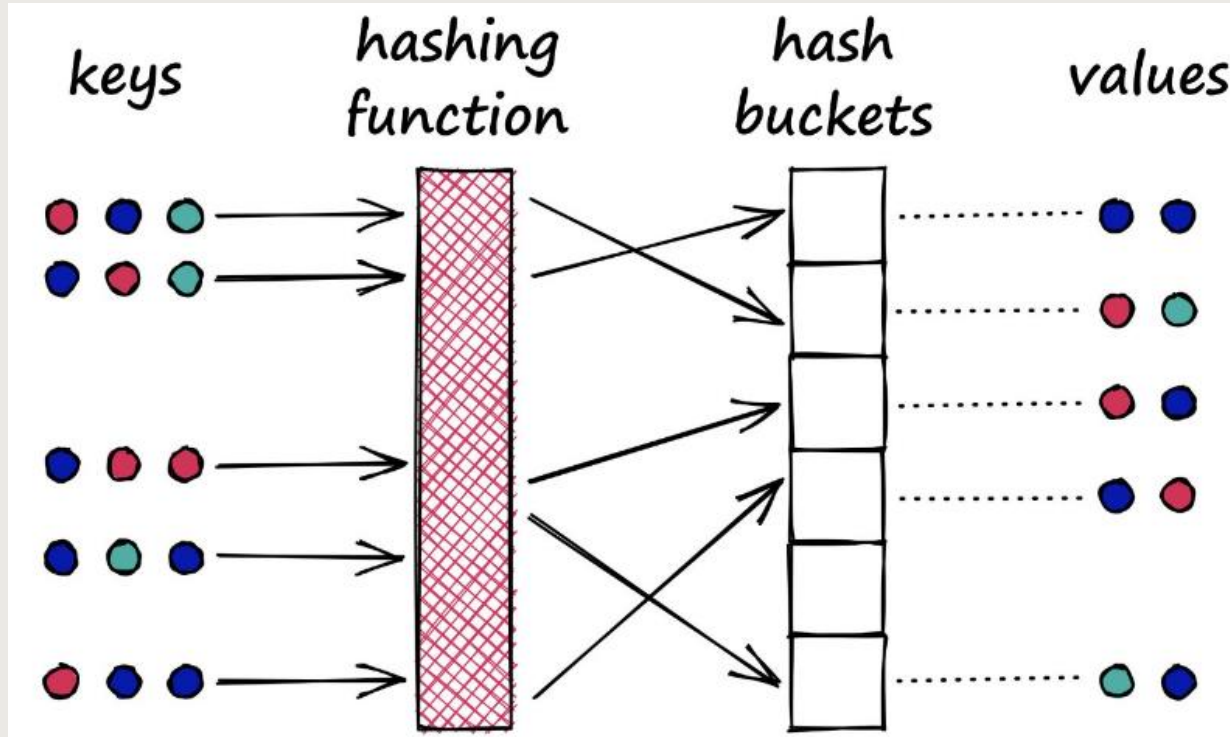
NetVLAD (Net of VLAD)



CNN architecture with the NetVLAD layer



Hash based indexing

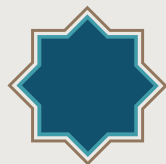


Hash based indexing



Kỹ thuật băm Locality sensitive hashing(LSH):

- ☐ Tất cả các vector đặc trưng trong cơ sở dữ liệu mà có cùng giá trị đầu ra sau khi trải qua một hàm băm h thì sẽ được đặt vào cùng một ô.
- ☐ Khi truy vấn ảnh, áp dụng hàm băm để trả ra kết quả và ánh xạ tới 1 ô đã có sẵn trong database.



Learned Secondary Indexing



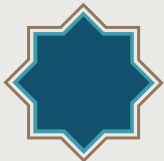
Look up



Learned Index



Fingerprint Vector	10	01	10	00	11	10	01	00	10	01
Permutation Vector	6	9	8	5	4	0	3	2	7	1
Base data	19	42	31	21	15	12	1	34	7	5



Learned Secondary Indexing



Gồm 3 phần: permutation vector, learned index, fingerprint vector

- Learned index: Ánh xạ Lookup key đến bounded search range
- Permutation vector: nén base data bằng cách thiết lập phần tử thứ i của tập dữ liệu d là $d[p[i]]$.
- Fingerprint vector: lưu mỗi giá trị một kích thước cố định, nếu fingerprint trong lookup có trong fingerprint vector thì tiếp tục truy cập, ngược lại thì dừng truy cập để tiết kiệm bộ nhớ.



Độ đo



Cho hai vector khác không a, b

Euclid	Cosin	Mahalanobis
$d_e = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$	$d_c = \cos(\theta) = \frac{a \cdot b}{\ a\ \ b\ }$ $= \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}$	$d_m = \sqrt{(a_i - b_j)^T c^{-1} (a_i - b_j)}$ c-1: covariance matrix



Độ đo

Ví dụ



Ngân

[1, 0, 0]



Minh

[0, 1, 0]



Nhung

[0, 0, 1]



?

[0.9, 0.07, 0.03]



?

[0.02, 0.96, 0.04]

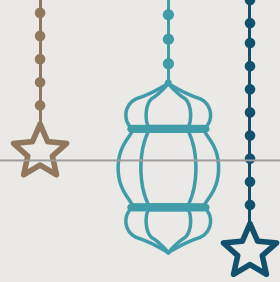


?

[0.1, 0.85, 0.05]



Độ đo



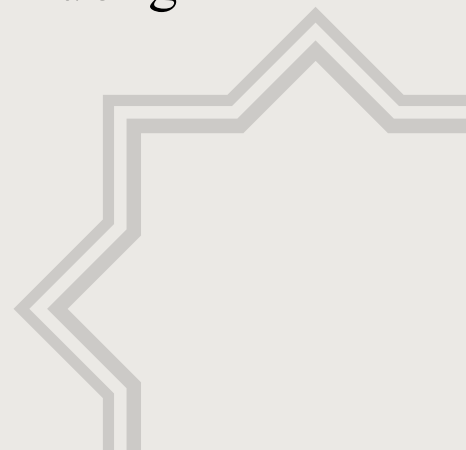
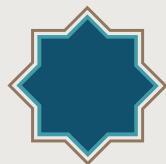
Euclid	$d_e = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$ $= \sqrt{(0.9 - 1)^2 + (0.07 - 0)^2 + (0.03 - 0)^2}$ $= 0.126$
Cosin	$d_c = \frac{\sum_{i=1}^n a_i b_i}{\sqrt{\sum_{i=1}^n a_i^2} \sqrt{\sum_{i=1}^n b_i^2}}$ $= \frac{0.9 \times 1 + 0.07 \times 0 + 0.03 \times 0}{\sqrt{0.9^2 + 0.07^2 + 0.03^2} \sqrt{1^2 + 0^2 + 0^2}}$ $= 0.996$
Mahalanobis	$\text{Covariance matrix } c^{-1} = \begin{bmatrix} 0.2412 & 0.2833 & 9.6667 \\ 0 & 0.3333 & 12.6667 \\ 0 & 0 & 6.7111 \end{bmatrix}$ $d_m = \sqrt{[0.9 \quad 0.07 \quad 0.03] \times \begin{bmatrix} 0.2412 & 0.2833 & 9.6667 \\ 0 & 0.3333 & 12.6667 \\ 0 & 0 & 6.7111 \end{bmatrix} \times \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}}$ $= \sqrt{0.0058} = 0.08$



Độ đo

Nhận xét:

- Khoảng cách euclid và mahalanobis có giá trị nhỏ \Rightarrow hình ảnh có độ tương đồng lớn.
- Cosin ra kết quả ngược lại (giá trị lớn gần bằng một) nhưng theo ý nghĩa của độ đo cosin thì đó là một kết quả tốt (khá tương đồng với query).



Liệu các độ đo bình thường có hiệu quả?



Tuy nhiên, **không phải tất cả** các khoảng cách **đều phù hợp** cho tác vụ CBIR. Trong một số trường hợp, khoảng cách Euclid không đủ tốt để phân biệt các ảnh tương đồng và không tương đồng. Từ đó, **distance metric learning** được sử dụng kết hợp với các độ đo để học một khoảng cách phù hợp cho tác vụ CBIR.

Giả sử: tập dữ liệu ảnh chứa các bức ảnh của động vật, bao gồm cả **hổ và gấu**. Nếu bạn có hai bức ảnh của **2 loài gấu khác nhau**, chúng rất giống nhau về hình dạng, thì khoảng cách Euclid giữa chúng có thể khá nhỏ, dù chúng có thể khác nhau về màu sắc và một số đặc điểm khác. Nếu chỉ dùng khoảng cách euclid thì có thể chưa chính xác.



Triplet network

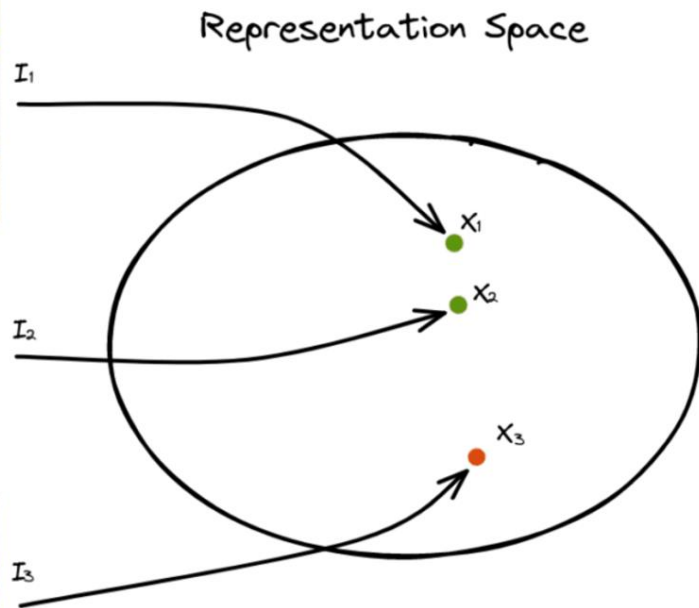
anchor



positive



negative



Triplet network



Vấn đề cần giải quyết:

- Minimize khoảng cách giữa anchor với positive?
- Maximize khoảng cách giữa anchor với negative?



Triplet network

Bộ ba ta có ảnh anchor (a), positive (p) và negative (n):
(a,p,n) chia thành các cặp: (a,p); (a,n)

Hàm triplet loss function:

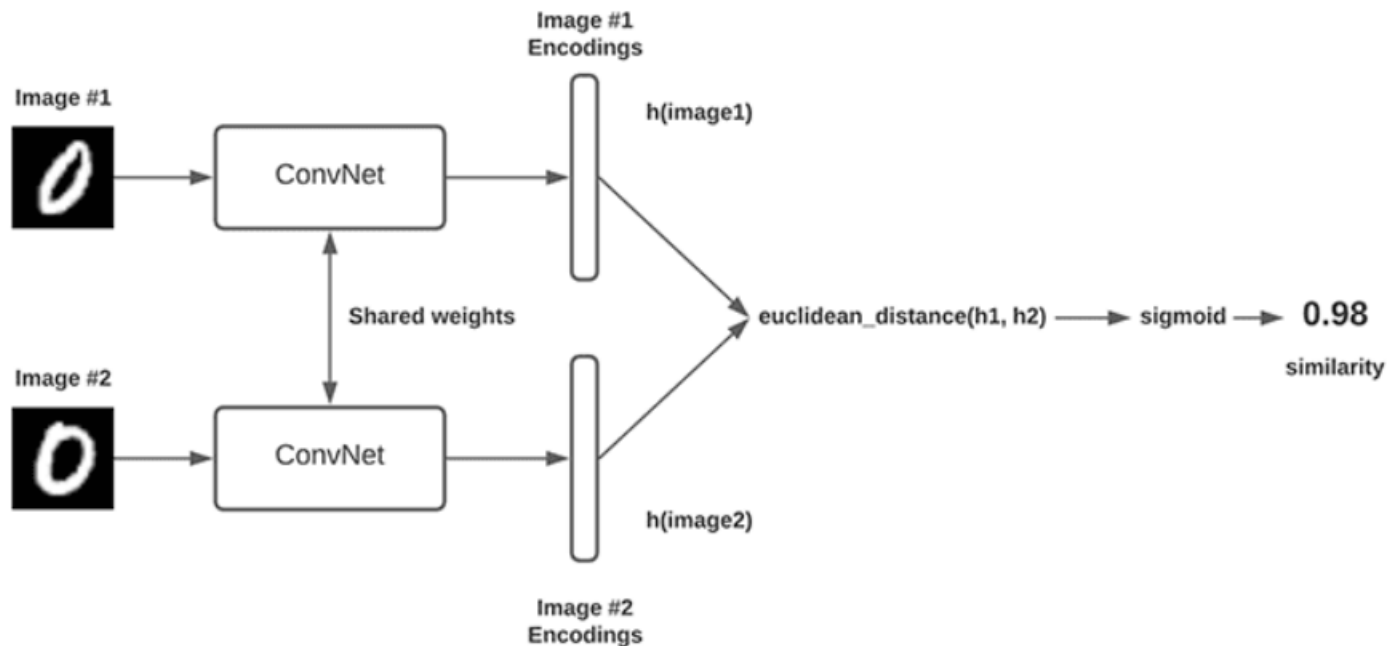
$$L = \max(d(a,n) - d(a,p) + \text{margin}, 0)$$

margin là một **hằng số** (thường được đặt là 1), đảm bảo khoảng cách giữa anchor và negative **lớn hơn** khoảng cách giữa anchor và positive một lượng tối thiểu.

Tối ưu hóa gradient descent: stochastic gradient descent (SGD)



Siamese network



Siamese network



- Input: hai ảnh x_i, x_j
- Mỗi ảnh sẽ đi qua ConvNet của Siamese Network để xử lý, sau đó mỗi ConvNet sẽ trả về một vector (h_1 và h_2)
- Tính khoảng cách Euclidean giữa h_1 và h_2 .
- Dùng hàm Sigmoid để chuẩn hóa giá trị khoảng cách Euclidean về khoảng $[0, 1]$. Kết quả của hàm Sigmoid càng gần 1 thì hai hình ảnh càng giống nhau.



Siamese network

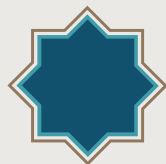


Siamese loss function

$$\begin{aligned} L_{Siame}(x_i, x_j) \\ = \frac{1}{2} s(x_i, x_j) D(x_i, x_j) + \frac{1}{2} (1 - s(x_i, x_j)) \max(0, m - D(x_i, x_j)) \end{aligned}$$

$S(x_i, x_j)$: Sigmoid function

$D(x_i, x_j)$: khoảng cách Euclid



Demo

Dataset

CBIR_dataset


[Data Card](#) [Code \(4\)](#) [Discussion \(0\)](#)

▲ 20


New Notebook

Download (243 MB)


dataset (4738 files)




0.jpg
50.64 kB




1.jpg
49.7 kB




10.jpg
49.12 kB




100.jpg
50.41 kB




1000.jpg
35.18 kB




1001.jpg
61 kB




1002.jpg
56.33 kB




1003.jpg
61.48 kB




1004.jpg
41.91 kB




1005.jpg
47.37 kB




1006.jpg




1007.jpg



1008.jpg



1009.jpg



101.jpg

Data Explorer

Version 1 (242.96 MB)

dataset

0.jpg

1.jpg

10.jpg

100.jpg

1000.jpg

1001.jpg

1002.jpg

1003.jpg

1004.jpg

1005.jpg

1006.jpg

1007.jpg

1008.jpg

1009.jpg

101.jpg

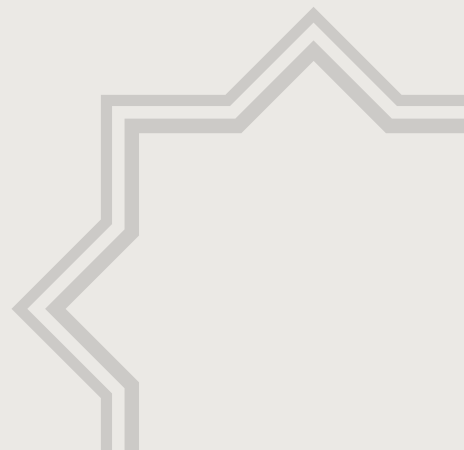
1010.jpg

1011.jpg

1012.jpg

Demo

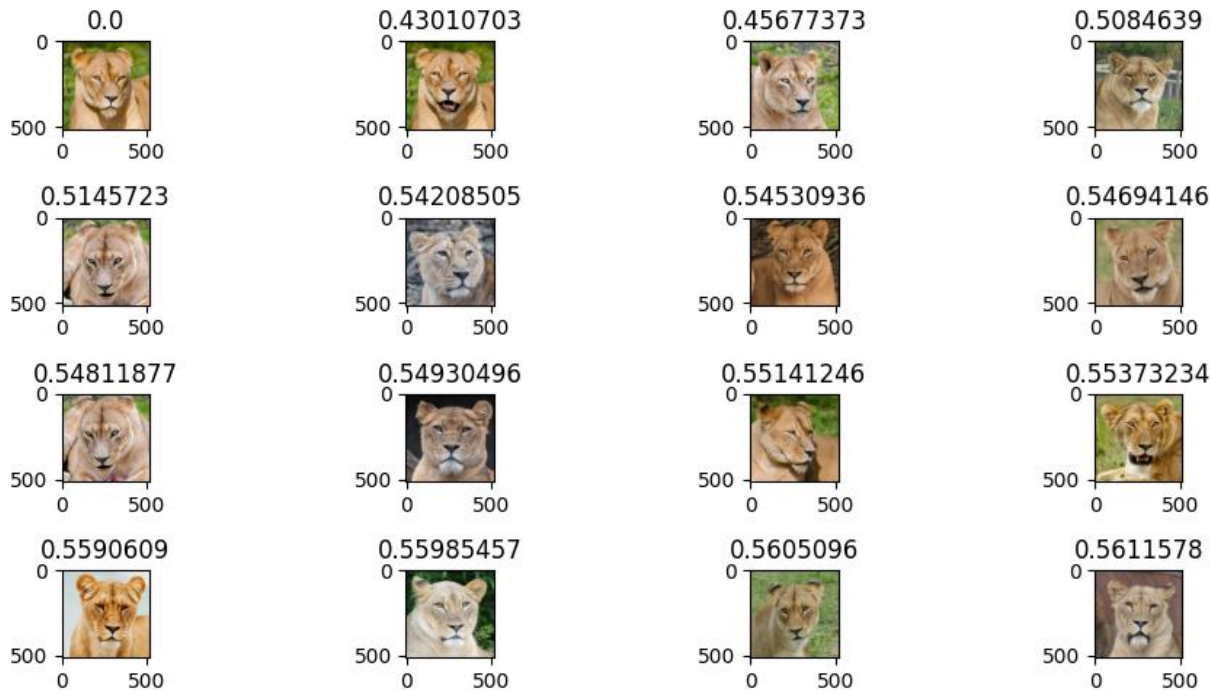
Query: ảnh 4571 trong dataset



Demo

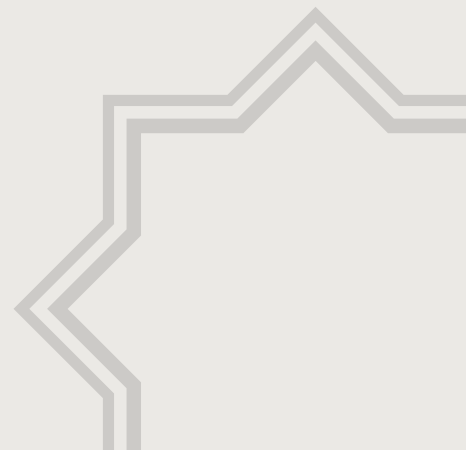


Output:

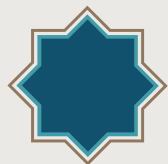
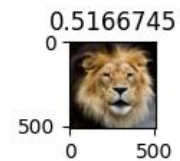
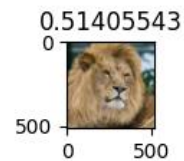
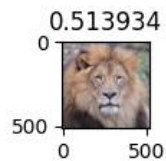
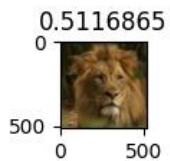
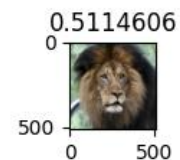
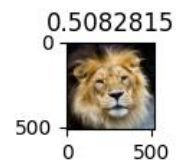
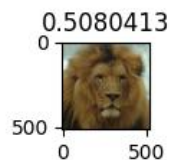
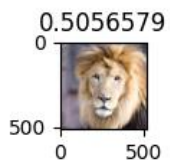
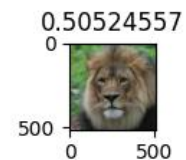
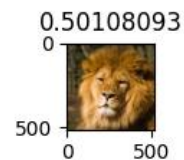
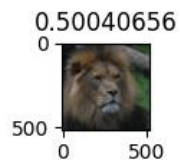
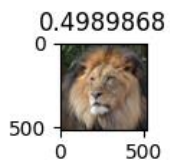
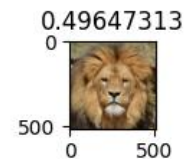
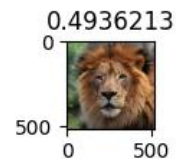
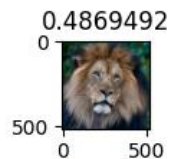
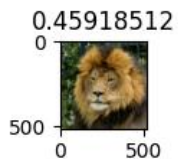


Demo

Query: ảnh download từ internet



Demo



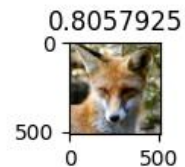
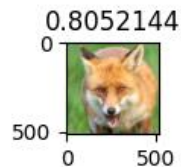
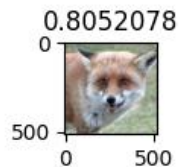
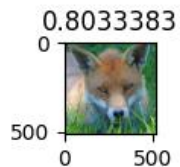
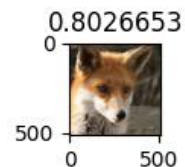
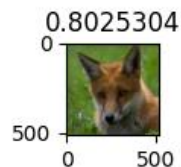
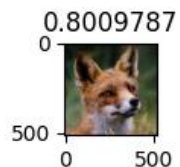
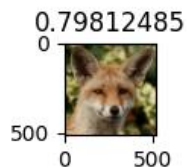
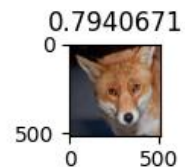
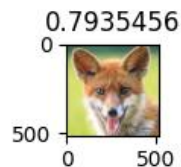
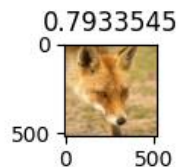
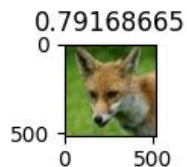
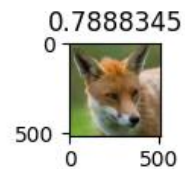
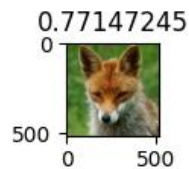
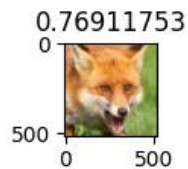
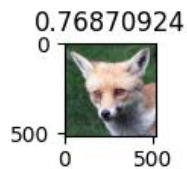
Demo



Query: ảnh download từ internet

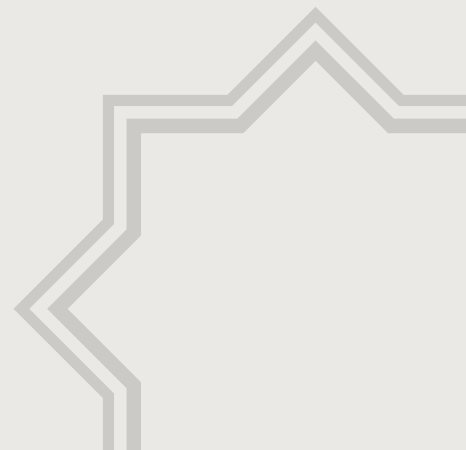


Demo

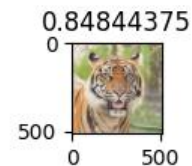
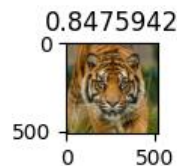
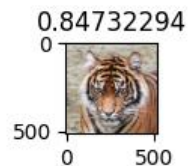
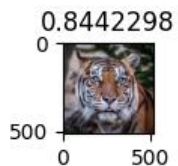
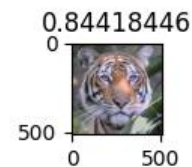
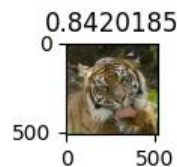
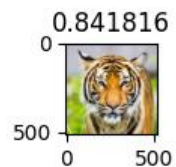
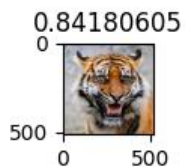
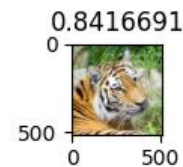
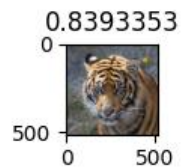
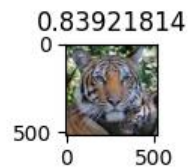
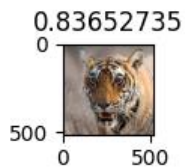
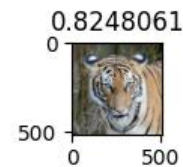
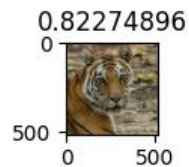
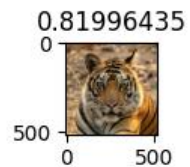
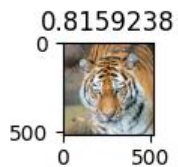


Demo

Query: ảnh download từ internet



Demo



**Thanks for
your
attention!**

التاريخ

