

TIẾP CẬN MÔ HÌNH HỌC CHUYÊN GIA CHO BÀI TOÁN TRUY VẤN HÌNH ẢNH DỰA TRÊN NỘI DUNG

Trần Văn Khánh, Huỳnh Thị Kim Chi
Trường Đại học Khánh Hoà

Thông tin chung:

Ngày nhận bài: 16/10/2023

Ngày phản biện: 20/10/2023

Ngày duyệt đăng: 05/11/2023

Title:

Transfer learning model approach for content-based image retrieval problem.

Từ khóa:

truy vấn hình ảnh, độ tương đồng hình ảnh, trích xuất đặc trưng, mạng nơ ron tích chập.

Keywords:

Image retrieval, image similarity, feature extraction, CNN.

TÓM TẮT: Trong bài báo, tác giả đề xuất phương pháp trích xuất đặc trưng hình ảnh sử dụng mô hình mạng nơ ron tích chập cho bài toán truy vấn hình ảnh dựa trên nội dung (Content-Based Image Retrieval - CBIR). Mục đích của đề xuất là giảm khoảng cách ngữ nghĩa giữa các tính năng cấp thấp và tính năng cấp cao, từ đó cải thiện kết quả của việc truy vấn hình ảnh. Mô hình mạng DenseNet 121 kết hợp kỹ thuật học chuyển giao thực thi việc học cách trích xuất đặc trưng từ tập cơ sở dữ liệu và sử dụng kiến thức học được cho việc trích xuất đặc trưng hình ảnh truy vấn. Thử nghiệm trên tập dữ liệu Corel, phương pháp đề xuất của tác giả được so sánh với các công trình công bố gần đây, đánh giá tính hiệu quả và áp dụng được trong các hệ tìm kiếm hình ảnh.

ABSTRACT: In the paper, the author proposes an image feature extraction method using a convolutional neural network model for the problem of content-based image retrieval (Content-Based Image Retrieval - CBIR). The purpose of the proposal is to reduce the semantic gap between low-level features and high-level features, thereby improving the results of image querying. The DenseNet 121 network model combines transfer learning techniques that implement learning to extract features from a database and using the learned knowledge for query image feature extraction. Experimenting on the Corel dataset, the author's proposed method is compared with recently published works, evaluating its effectiveness and applicability in image retrieval systems.

1. Giới thiệu

Sự phát triển nhanh chóng trong việc thu thập hình ảnh và công nghệ thông tin đã tạo ra một lượng lớn các tập dữ liệu hình ảnh. Do đó, việc quan trọng là làm như thế nào để tạo ra được một hệ thống xử lý hiệu quả các tập dữ liệu lớn, đặc biệt là các giải pháp tìm kiếm nhanh một hình ảnh trong tập dữ liệu lớn hay còn gọi là truy vấn hình ảnh dựa trên nội dung. Phương pháp này được khuyến khích dùng để vượt qua những nhược điểm của phương thức truy vấn ảnh dựa trên văn bản [14].

CBIR được sử dụng để tìm kiếm những hình ảnh có nội dung tương tự tập từ tập dữ liệu hình ảnh với một hình ảnh truy vấn cụ thể [15], [16]. CBIR là một bước nền tảng cho nhiều ứng dụng và có thể được dùng để xử lý một lượng lớn các công việc. Ví dụ, khi tìm kiếm trên các trang web hoặc trong một cơ sở dữ liệu hình ảnh lớn, người dùng có thể gặp khó khăn trong việc thể hiện mong muốn của mình. Với CBIR, điều đó được gọi là khoảng cách ngữ nghĩa và nó có thể được giải quyết bằng cách cung cấp hình ảnh mẫu thay vì mô tả bằng văn bản [11], [12].

Nhìn chung, một hệ thống CBIR bao gồm một số giai đoạn bắt buộc và một số giai đoạn tùy chọn. Giai đoạn đầu tiên trong CBIR là đưa hình ảnh truy vấn vào bởi người dùng. Tất cả quá trình xử lý được áp dụng cho hình ảnh truy vấn đều được áp dụng cho tất cả hình ảnh trong tập cơ sở dữ liệu và cùng một thứ tự [13]. Thông thường, quá trình này được thực hiện trên ảnh truy vấn khi người dùng gửi và được gọi là quá trình trực tuyến; Quá trình này cũng được thực hiện trên tập cơ sở dữ liệu trước đó và được gọi là quá trình ngoại tuyến. Một quá trình xử lý tùy chọn có thể bao gồm trong hệ thống CBIR là tiền xử lý, quá trình có thể bao gồm thay đổi kích thước, phân đoạn, xóa nhiễu, v.v. Giai đoạn trích xuất thông tin được thực hiện sau giai đoạn tùy chọn trên, đây là giai đoạn quan trọng nhất giúp chuyển đổi khái niệm trực quan thành dạng số. Các đặc trưng được trích xuất có thể ở dạng đặc trưng cấp thấp (màu sắc, hình dạng, cấu trúc, thông tin không gian). Một giai đoạn tùy chọn khác sau trích xuất đặc trưng là chuẩn hoá hoặc phân lớp. Giai đoạn cuối cùng là đo độ tương đồng giữa đặc trưng trích xuất của hình ảnh truy vấn và tất cả hình ảnh trong tập cơ sở dữ liệu để truy vấn những hình ảnh liên quan nhất. Phản hồi liên quan là một giai đoạn khả thi giúp nâng cao độ chính xác kết quả thông qua sự can thiệp của con người bằng cách quyết định hình ảnh trả về có liên quan hay không liên quan.

2. Công trình liên quan

Năm 2015, Shrivastava [1] đã đề xuất một hệ thống CBIR với một cấu trúc truyền thẳng gồm 3 giai đoạn để truy vấn hình ảnh. Giai đoạn đầu tiên, dựa vào đặc trưng màu sắc lấy ra N hình ảnh phù hợp từ tập cơ sở dữ liệu gồm M hình ảnh. Tác giả sử dụng biểu đồ màu để tính toán đặc trưng màu sắc. Giai đoạn 2, lấy ra P hình ảnh liên quan từ N hình ảnh (ở giai đoạn 1) dựa trên đặc trưng kết cấu sử dụng bộ lọc Gabor. Giai đoạn cuối cùng, sử dụng bộ mô tả Fourier để tính toán đặc

trung hình dạng và truy vấn lấy ra K hình ảnh liên quan từ P hình ảnh (ở giai đoạn 2). Theo phương pháp này, phản hồi liên quan (Relevance feedback) có thể được áp dụng để thay đổi các giá trị N, K, P để cải thiện độ chính xác. Hệ thống được kiểm tra trên hai tập dữ liệu Corel và CIFAR với độ chính xác tương ứng là 0.7690 và 0.8590. Tuy nhiên, hệ thống đề xuất chưa sử dụng giai đoạn phân loại hình ảnh dựa trên thông tin không gian.

Năm 2016, Anandh [2] đã đưa ra một hệ thống CBIR dựa trên sự kết hợp của màu sắc, kết cấu và hình dạng. Tương quan tự động màu sắc, biến đổi Gabor, biến đổi Wavelet đã được sử dụng để trích xuất màu sắc, hình dạng và kết cấu. Nhóm tác giả đã sử dụng khoảng cách Manhattan để tính toán độ tương đồng giữa hình ảnh truy vấn và hình ảnh trong tập cơ sở dữ liệu. Độ chính xác trung bình thu được tương ứng 0.8300, 0.8800 và 0.7000 lần lượt cho các tập cơ sở dữ liệu Corel, Li và Caltech101. Hạn chế chính của hệ thống là độ phức tạp tính toán tăng do kết hợp nhiều đặc trưng.

Năm 2018, Sajjad [3] đề xuất một hệ thống CBIR bất biến với xoay kết cấu và thay đổi màu sắc. Hệ thống đề xuất dựa trên việc nổi đặc trưng màu sắc với đặc trưng kết cấu để tạo thành véc-tơ đặc trưng với kích thước 360. Để trích xuất đặc trưng màu sắc, hình ảnh được chuyển đổi về không gian màu HSV và lượng tử hoá thông qua biểu đồ màu. Để bất biến với những thay đổi độ sáng, tác giả chỉ sử dụng kênh màu HUE và Saturation. Mẫu nhị phân cục bộ xoay (rotated local binary pattern – RLBP) được dùng để trích xuất đặc trưng kết cấu bất biến xoay. Hệ thống đề xuất được đánh giá thông qua thực nghiệm trên corel với độ chính xác là 0.8777.

Năm 2019, Sezavar [16] đã đề xuất tiếp cận truy vấn hình ảnh dựa trên nội dung sử dụng CNN để trích xuất đặc trưng cấp cao. Các lớp cuối của AlexNet [17] được dùng để trích xuất đặc trưng do lớp cuối của nó có véc-tơ đặc trưng nhỏ nên giảm được chi phí

tính toán. Đề xuất này được kiểm thử trên các tập dữ liệu ALOI, Corel và MPEG-7, kết quả thu được tốc độ truy vấn tốt. Mặc dù vậy, đại diện không gian nhỏ nên thu được độ chính xác thấp hơn so với các hiện nghiệm cùng thời điểm.

Zheng [4] đã đề xuất hệ thống CBIR hoàn thiện dựa trên VGGNet, tác giả sử dụng CNN để trích xuất đặc trưng, sử dụng tập dữ liệu trường hấp dẫn và nhãn điểm tương đồng thay cho việc dùng nhãn thông thường. Hệ thống đề xuất thu được độ chính xác 0.9620, 0.9410, 0.8850 trên các tập dữ liệu tương ứng là Oxford Paris, Holidays và Caltech 101. Tuy nhiên, hệ thống mất nhiều thời gian để xây dựng tập dữ liệu trường hấp dẫn và cần cải thiện tốc độ hơn trong quá trình huấn luyện và kiểm thử. Với đề xuất này tác giả chưa thực nghiệm trên tập dữ liệu Corel.

Năm 2021, bằng cách kết hợp nhiều kỹ thuật học máy như k-NN, CNN hoặc kết hợp kỹ thuật học máy với một cấu trúc dữ liệu và các nhóm đặc trưng hình ảnh trên các bộ ảnh COREL, Wang, CIFAR-100, Caltech101, Nhi và cộng sự [5] đã thực hiện, phương pháp phân lớp hình ảnh bằng cách kết hợp cấu trúc cây C-Tree và thuật toán phân lớp k-NN tại nút lá trên C-Tree đã thu được kết quả khả quan. Trong công trình này, các tác giả đã thực nghiệm trên bộ ảnh Corel và Wang với hiệu suất phân lớp là 0,7302 và 0,6894.

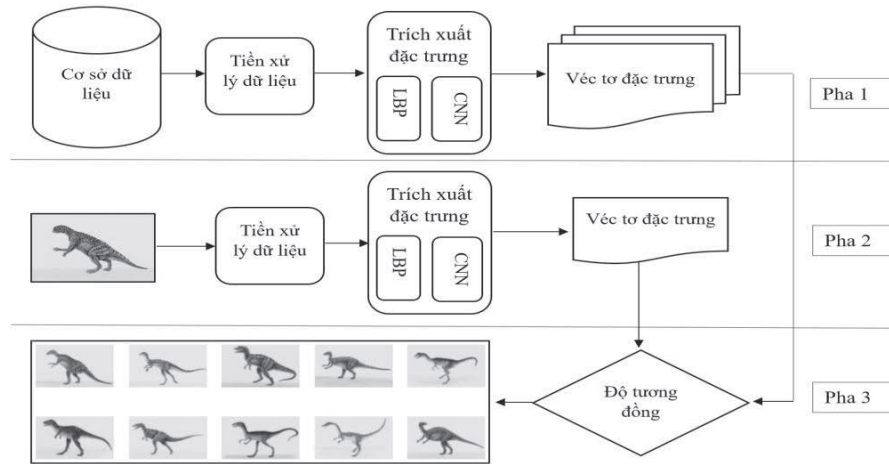
Từ phân tích các công trình liên quan cho thấy mô hình truy vấn ảnh dựa trên véc-tơ đặc trưng được đánh giá là khả thi và hiệu quả thời gian gần đây. Trong bài báo này, chúng tôi tiếp cận theo phương pháp trích xuất véc-tơ đặc trưng theo các kỹ thuật khác nhau. Từ đó, chúng tôi thực nghiệm trên bộ dữ liệu ảnh Corel để minh chứng cho mô hình tìm kiếm ảnh tương tự dựa trên kỹ thuật trích xuất véc-tơ đặc trưng kết hợp kết cấu hình ảnh là hiệu quả.

Đóng góp của bài báo gồm: (1) Trích xuất đặc trưng kết cấu của hình ảnh dựa trên mẫu nhị phân cục bộ; (2) Sử dụng mô hình mạng nơ ron tích chập trích xuất đặc trưng cấp cao để tạo véc-tơ đặc trưng cho hình ảnh; (3) Đề xuất mô hình tìm kiếm ảnh tương tự dựa trên véc-tơ đặc trưng đã trích xuất; (4) Xây dựng thực nghiệm tìm kiếm ảnh tương tự dựa trên véc-tơ đặc trưng và so sánh với một số công trình cùng bộ dữ liệu.

3. Mô hình đề xuất

Tác giả sử dụng mô hình mạng học chuyên gia DenseNet 121 với bộ trọng số đã được huấn luyện trên tập dữ liệu ImageNet [9] để trích xuất thông tin đặc trưng mức cao, kết hợp với mẫu nhị phân cục bộ (Local binary pattern – LBP) để trích xuất kết cấu hình ảnh. LBP cung cấp cấu trúc thông tin của các điểm ảnh mức độ xám. Tính toán của LBP đơn giản, nó không nhạy cảm với sự thay đổi của mức độ xám mà nhạy cảm với nhiễu. Việc kết hợp đặc trưng này giúp quá trình trích xuất thu được chi tiết các đặc trưng. Véc-tơ đặc trưng được nối từ kết quả của hai quá trình trên có độ dài 50186 (với trích xuất DenseNet 121 ta thu được véc-tơ có độ dài 50176, LBP thu được véc-tơ có độ dài 10). Mô hình đề xuất sau đây được gọi là DeLBP-CBIR.

Mô hình hoạt động dựa trên các giai đoạn khác nhau như trích xuất đặc trưng, so sánh độ tương đồng. Ở pha đầu tiên, mô hình DenseNet 121 được khởi tạo để trích xuất đặc trưng cho hình ảnh trong tập dữ liệu huấn luyện. Pha thứ hai, mô hình thực hiện việc trích xuất đặc trưng hình ảnh truy vấn. Pha cuối, hệ thống thực hiện việc so sánh độ tương đồng giữa véc-tơ đặc trưng của hình ảnh truy vấn và các véc-tơ đặc trưng của hình ảnh trong tập dữ liệu để cho ra kết quả các hình ảnh có độ tương đồng với hình ảnh truy vấn (Hình 1).



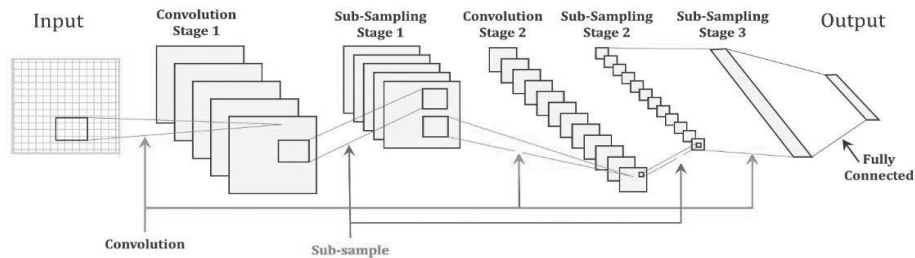
Hình 1. Quá trình xử lý của mô hình truy vấn hình ảnh dựa trên nội dung.

3.1. Học sâu (Deep learning - DL)

DL là một nhánh nhỏ của học máy, nhận được rất nhiều quan tâm thời gian gần đây trong việc xử lý các bài toán thực tế. Cấu trúc của DL chứa các thuật toán học máy được thiết kế từ việc lấy cảm hứng của bộ não con người. Những thuật toán được tổ chức và vận dụng thông tin bằng cách đưa chúng qua các giai đoạn thể hiện và chuyển đổi. Sự thành công của những thuật toán học sâu đã tạo cảm hứng cho việc ứng dụng nó vào hệ thống CBIR để làm giảm khoảng cách ngữ nghĩa. Cấu trúc của DL làm nó có khả năng ánh xạ dữ liệu của lớp đầu vào đến lớp đầu ra mà không phụ thuộc những đặc trưng con người cung cấp [6].

Thuật toán DL bao gồm mạng nơ ron tích chập (Convolutional neural network – CNN), mạng nơ ron sâu (Deep neural network –

DNN), mạng tin sâu (Deep belief network) và học máy Boltzmann. CNN đã thể hiện được hiệu suất vượt trội trong các ứng dụng thị giác máy tính như nhận diện khuôn mặt, phát hiện đối tượng và phân đoạn ngữ nghĩa [7] và đặc biệt trong hệ thống CBIR. CNN chứa ba loại lớp: lớp tích chập, lớp gộp và lớp kết nối đầy đủ (Hình 2). Các bộ lọc được áp dụng cho hình ảnh đầu vào thông qua các lớp tích chập để học các đặc trưng trong khi đó chức năng của lớp trung gian (lớp gộp) là làm giảm độ lớn của đầu vào. Lớp kết nối đầy đủ dự đoán nhãn của hình ảnh đầu vào. CNN bất biến với phép dịch, thay đổi tỷ lệ và phép quay làm cho nó trở nên có giá trị đối với các ứng dụng thị giác máy tính. CNN dựa trên dữ liệu đã được gán nhãn [7], đây được xem là một hạn chế của nó. Điểm khác biệt nữa là CNN không yêu cầu trích xuất tính năng thủ công [8].



Hình 2. Cấu trúc chung của một CNN.

3.2. Mô hình DenseNet 121

Trong bài báo này, mô hình DenseNet 121 được áp dụng cho việc trích xuất đặc trưng hình ảnh trong tập cơ sở dữ liệu và hình ảnh truy vấn. Mô hình DenseNet 121 chứa 121 lớp,

bao gồm một số thành phần giống như phiên bản DenseNet khác: Các lớp bottleneck, các lớp chuyển đổi và các khối dense:

Lớp bottleneck: lớp này giúp giảm tham số của các bản đồ đặc trưng (feature maps) trước khi áp dụng các phép tích chập kích thước lớn.

Bottleneck có thể được biểu diễn như sau:

$$\text{Bot}(x) = \text{Re}(\text{BN}(\text{Conv}1 \times 1(x)))$$

Khối dày đặc (dense block) và lớp chuyển đổi (transition layer): mỗi khối dày đặc gồm nhiều lớp liên kết mật thiết với nhau, các khối dày đặc thường kết thúc bằng các lớp chuyển đổi để giảm kích thước của bản đồ đặc trưng và kiểm soát số lượng bản đồ đặc trưng được truyền đến khối dày đặc tiếp theo. Lớp chuyển đổi có thể được biểu diễn như sau:

$$\text{Tran}(x) = \text{Re}(\text{BN}(\text{Conv}1 \times 1(x)))$$

$$\text{Tran}(x) = \text{AvgP}(x)$$

Trong đó:

x : là đầu vào của lớp

$\text{Conv}1 \times 1(x)$: thể hiện phép tích chập 1×1

BN : đại diện cho quá trình chuẩn hóa dữ liệu

Re : làm hàm kích hoạt Rectified Linear Unit

AvgP : là lớp áp dụng phép gộp trung bình

Kết nối dày đặc (Dense connective): mỗi lớp trong khối dày đặc nhận đầu vào từ tất cả các lớp trước nó. Điều này có thể được biểu diễn như sau:

$$x_k = H_k([x_0, x_1, \dots, x_{k-1}])$$

Trong đó:

x_k : là đầu ra của lớp thứ k

H_k : đại diện cho hàm kích hoạt của lớp thứ k

$[x_0, x_1, \dots, x_{k-1}]$: là sự kết hợp của các bản đồ đặc trưng từ tất cả các lớp trước đó đến $k - 1$

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 64$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$	$\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$
Classification Layer	1×1	7×7 global average pool			
		1000D fully-connected, softmax			

Hình 3. Cấu trúc của mô hình DenseNet.

3.3. Mẫu nhị phân cục bộ (LBP)

LBP là thuật toán mô tả cấu trúc cục bộ của hình ảnh bằng cách so sánh mỗi điểm ảnh (pixel) với các điểm ảnh xung quanh. Thuật toán được phát triển để trích xuất thông tin trội, giúp cho bài toán phân lớp đạt độ chính xác cao hơn. Với một điểm ảnh trung tâm có tọa độ (x_c, y_c) , một tập nhị phân có thứ tự được xác định là LBP thu được bằng cách so sánh giá trị màu xám của điểm ảnh trung tâm (x_c, y_c) với tám điểm ảnh xung quanh.

$$\text{LBP}(x_c, y_c) = \sum_{n=0}^7 S(i_n - i_c) 2^n$$

Với i_c thể hiện giá trị của điểm ảnh trung tâm, i_n lần lượt là giá trị của 8 điểm ảnh xung quanh. LBP được chứng minh là bất biến với bất kì chuyển đổi mức xám đơn điệu nào và mã nhị phân lân cận giữ nguyên sau khi chuyển đổi.

$$S(i_n - i_c) = \begin{cases} 1 & i_n - i_c \geq 0 \\ 0 & i_n - i_c < 0 \end{cases}$$

Hình 4 là ví dụ minh họa cách tính LBP với ngưỡng xử lý là 8 điểm ảnh xung quanh. Hình 4a) đưa ra một vùng ảnh gồm 9 điểm ảnh. Hình 4b) dựa vào công thức (2) so sánh giá trị của điểm ảnh trung tâm và các điểm ảnh xung

quanh, cho ra kết quả S tại điểm ảnh là 1 hoặc 0. Hình 4c) dựa vào công thức (1) ta tính được:

$$LBP = 0*2^0 + 0*2^1 + 1*2^2 + 0*2^3 + 1*2^4 + 1*2^5 + 1*2^6 + 0*2^7 = 116.$$

78	89	145	0	0	1	1	2	4
91	98	143	0		1	8		16
99	123	34	1	1	0	32	64	128

a) Giá trị điểm ảnh mức xám

b) Tập giá trị nhị phân

c) Tập giá trị 2^n

Hình 4. Minh hoạ mẫu nhị phân cục bộ

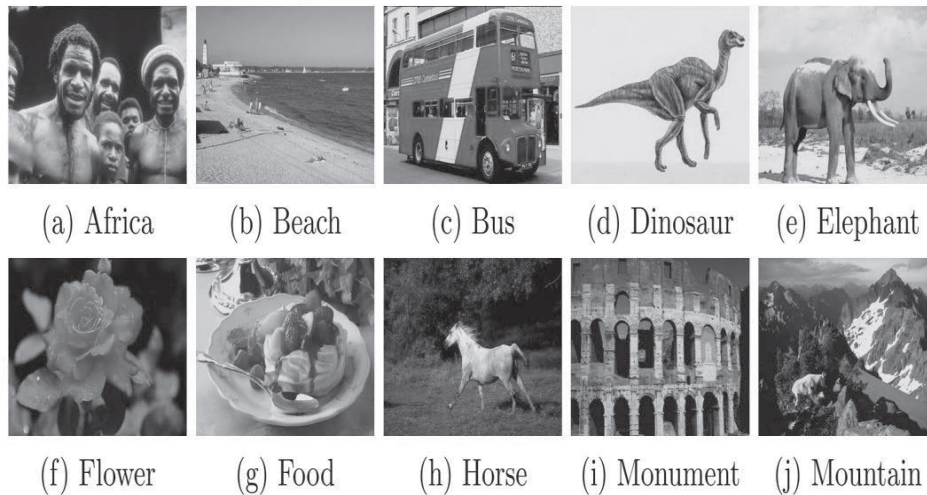
4. Thực nghiệm

4.1. Cài đặt thực nghiệm

Hệ thống đề xuất được thực nghiệm trên bộ dữ liệu Corel [10]. Mỗi ảnh được trích xuất thành một véc-tơ đặc trưng với độ lớn 50186 và thuộc một phân lớp. Thực nghiệm được

thực hiện dựa trên ngôn ngữ lập trình Python. Cấu hình máy tính thực nghiệm: Intel(R) CoreTM i5-5200U, CPU 2.2GHz, Ram 16G và hệ điều hành Windows 11 Professional.

Tập cơ sở dữ liệu Corel gồm 1000 ảnh được chia thành 10 chủ đề (Hình 5): beach, bus, castle, dinosaur, elephant, flower, horse, meal, mountain, peoples. Trong thực nghiệm này, việc truy vấn thực hiện so sánh độ tương đồng dựa trên chỉ số tương đồng cosin giữa các véc-tơ đặc trưng. Hai ảnh được gọi là tương đồng nếu độ tương đồng nhỏ hơn một giá trị ngưỡng ϵ cho trước. Số lượng ảnh tương đồng được lấy dựa trên các ảnh được lấy TopK các ảnh truy vấn để tính độ chính xác (precision) và độ phủ (recall).



Hình 5. Minh hoạ tập cơ sở dữ liệu Corel.

4.2. Đánh giá hiệu suất

Các phương pháp thông thường nhất dùng để đánh giá hiệu suất của hệ thống CBIR là độ chính xác (Precision – P), độ phủ (Recal – R). Sử dụng ma trận nhầm lẫn (Bảng 1) ta có thể thu được độ chính xác và độ phủ dựa vào công thức (1) (2).

Bảng 1. Ma trận nhầm lẫn.

		Lớp dự đoán (Predicted Class)	
		True (T)	False (F)
Lớp thực tế (Actual Class)	True (T)	TP	FN
	False (F)	FP	TN

$$\text{Precision (P)} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall (R)} = \frac{TP}{TP+FN} \quad (2)$$

Trong đó: TP: True Positive; FN: False Negative; TN: True Negative; FP: False Positive.

4.3. Kết quả thực nghiệm

Kết quả thực nghiệm trên nhóm hình ảnh thuộc foods thu được độ chính xác 0.860, 0.855 và 0.840 tương ứng trên Top10, Top20 và Top30 (Bảng 2). Đối với nhóm hình ảnh thuộc horses thu được độ chính xác 1.000, 1.000. và 0.990 (Bảng 3). Trong đó nhóm foods có 6 hình ảnh đạt độ chính xác tuyệt

đôi, nhóm horses ở Top10 và Top20 tất cả hình ảnh đều có độ chính xác truy vấn tuyệt đối, điều này cho thấy việc trích xuất thông

tin bậc cao kết hợp kết cấu hình ảnh mang lại hiệu suất cao cho truy vấn hình ảnh dựa trên nội dung.

Bảng 2. Kết quả trên tập hình ảnh foods.

Hình ảnh	Top10		Top20		Top30	
	P	R	P	R	P	R
900.jpg	0.900	0.100	0.800	0.178	0.833	0.278
901.jpg	0.500	0.056	0.600	0.133	0.567	0.189
902.jpg	1.000	0.111	0.950	0.211	0.933	0.311
903.jpg	1.000	0.111	1.000	0.222	1.000	0.333
904.jpg	0.300	0.033	0.300	0.067	0.267	0.089
905.jpg	0.900	0.100	0.900	0.200	0.800	0.267
906.jpg	1.000	0.111	1.000	0.222	1.000	0.333
907.jpg	1.000	0.111	1.000	0.222	1.000	0.333
908.jpg	1.000	0.111	1.000	0.222	1.000	0.333
909.jpg	1.000	0.111	1.000	0.222	1.000	0.333
Trung bình	0.860	0.096	0.855	0.190	0.840	0.280

Bảng 3. Kết quả trên tập hình ảnh horses

Hình ảnh	Top10		Top20		Top30	
	P	R	P	R	P	R
700.jpg	1.000	0.111	1.000	0.222	1.000	0.333
701.jpg	1.000	0.111	1.000	0.222	1.000	0.333
702.jpg	1.000	0.111	1.000	0.222	1.000	0.333
703.jpg	1.000	0.111	1.000	0.222	1.000	0.333
704.jpg	1.000	0.111	1.000	0.222	0.933	0.311
705.jpg	1.000	0.111	1.000	0.222	0.967	0.322
706.jpg	1.000	0.111	1.000	0.222	1.000	0.333
707.jpg	1.000	0.111	1.000	0.222	1.000	0.333
708.jpg	1.000	0.111	1.000	0.222	1.000	0.333
709.jpg	1.000	0.111	1.000	0.222	1.000	0.333
Trung bình	1.000	0.111	1.000	0.222	0.990	0.330

Kết quả tổng hợp trung bình các nhóm hình ảnh và trung bình của toàn tập cơ sở dữ liệu Corel thực nghiệm trên đề xuất của tác giả thu được với 0.938, 0.934 và 0.924 tương ứng với Top10, Top20, Top30 (Bảng 4). Kết quả so sánh với một số nghiên cứu cùng lĩnh

vực (Bảng 5), với độ chính xác thu được **0.938** cho toàn bộ tập dữ liệu Corel, đây là một hiệu suất khá cao so với các nghiên cứu cùng lĩnh vực. Và với hiệu suất như vậy có thể sử dụng để triển khai ứng dụng trên thực tế.

Bảng 4. Kết quả trung bình theo nhóm hình ảnh

Nhóm hình ảnh	Top10		Top20		Top20	
	P	R	P	R	P	R
Beaches	0.720	0.080	0.720	0.160	0.697	0.232
Bus	1.000	0.111	1.000	0.222	1.000	0.333
Dinosaurs	1.000	0.111	1.000	0.222	1.000	0.333
Elephants	1.000	0.111	1.000	0.222	1.000	0.333
Flowers	1.000	0.111	1.000	0.222	1.000	0.333
Foods	0.860	0.096	0.855	0.190	0.840	0.280
Horses	1.000	0.111	1.000	0.222	0.990	0.330
Monuments	0.950	0.106	0.945	0.210	0.927	0.309
Mountains ...	1.000	0.111	0.995	0.221	0.993	0.331
People ...	0.850	0.094	0.825	0.183	0.797	0.266
Trung bình	0.938	0.104	0.934	0.207	0.924	0.308

Bảng 5. Kết quả đối sánh với nghiên cứu cùng lĩnh vực.

Nghiên cứu	Độ chính xác
Shrivastava [1]	0.769
Anandh [2]	0.830
Sajjad [3]	0.877
Nhi, Nguyen Thi Uyen [5]	0.730
DeLBP-CBIR	0.938

Kết quả minh họa truy vấn hình ảnh dựa trên hệ thống CBIR đề xuất cho flowers, dinosaurs và horses (Hình 6). Hình lớn ở trên là hình truy vấn, 20 hình ảnh nhỏ ở dưới là hình tương tự được đưa ra từ kết quả truy vấn.



Hình 6. Minh họa kết quả truy vấn trực quan.

5. Kết luận

Chúng tôi đã thực hiện việc trích xuất đặc trưng kết cấu của hình ảnh dựa trên các LBP kết hợp với đặc trưng mức cao từ DenseNet 121, từ đó áp dụng vào bài toán truy vấn hình ảnh. Việc trích xuất đặc trưng đã thực hiện trên các bộ ảnh thông dụng Corel, đồng thời áp dụng tìm kiếm ảnh tương đồng dựa trên mô hình đã đề xuất.

Theo thực nghiệm cho thấy, phương pháp trích xuất đặc trưng áp dụng cho bài toán tìm kiếm ảnh là hiệu quả và có độ chính xác cao, cụ thể là 93.80% trên tập cơ sở dữ liệu Corel. Việc thực nghiệm trên bộ ảnh thông dụng Corel đã minh chứng tính khả thi của mô hình đề xuất và có thể áp dụng trên các hệ thống dữ liệu đa phương tiện.

Hướng nghiên cứu trong tương lai của chúng tôi là phân cụm bộ dữ liệu véc-tơ đặc trưng nhằm tăng tốc độ tìm kiếm hình ảnh tương đồng. Thay vì so sánh độ tương đồng của hình ảnh truy vấn với toàn bộ hình ảnh trong tập cơ sở dữ liệu, ta chỉ cần so sánh độ

tương đồng giữa hình ảnh truy vấn với cụm dữ liệu hình ảnh tương đồng.

Tài liệu tham khảo

- Shrivastava, N., & Tyagi, V. (2015, August). *An efficient technique for retrieval of color images in large databases*. *Computers & Electrical Engineering*, 46, 314–327. 10.1016/j.compeleceng.2014.11.009
- Ponomarev, A., Nalamwar, H. S., Babakov, I., Parkhi, C. S., & Buddhawar, G. (2016, February). *Content-based image retrieval using color, texture and shape features*. *Key Engineering Materials*, 685, 872–876. 10.4028/www.scientific.net/KEM.685.872
- Sajjad, M., Ullah, A., Ahmad, J., Abbas, N., Rho, S., & Baik, S. W. (2018, February). *Integrating salient colors with rotational invariant texture features for image representation in retrieval systems*.

- Multimedia Tools and Applications, 77(4), 4769–4789. 10.1007/s11042-017-5010-5
4. Zheng, Q., Tian, X., Yang, M., & Wang, H. (2019). *Differential learning: A powerful tool for interactive content-based image retrieval*. Engineering Letters, 27(1), 202–215. http://www.engineeringletters.com/issues_v27/issue_1/EL_27_1_23.pdf
 5. Nhi, Nguyen Thi Uyen, and Thanh Manh Le, (2021). *Semantic-Based Image Retrieval Using Balanced Clustering Tree*. World Conference on Information Systems and Technologies. Springer, Cham.
 6. Wan, J., *Deep learning for content-based image retrieval*. In Proceedings of the ACM International Conference on Multimedia - MM '14. New York, USA, 2014, pp. 157–166.
 7. Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). *Deep learning for computer vision: A brief review*. Computational Intelligence and Neuros
 8. Patil, A., & Rane, M. (2021). *Convolutional neural networks: an overview and its applications in pattern recognition*. In Smart Innovation, Systems and Technologies, 195(Insights into Imaging), 21–30. 10.1007/978-981-15-7078-0_3
 9. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei. (2009). *ImageNet: A Large-Scale Hierarchical Image Database*. IEEE Computer Vision and Pattern Recognition (CVPR).
 10. Kaggle, <https://www.kaggle.com/datasets/elkamel/co-rel-images>, truy cập 08/2023
 11. Jain M., & Singh S. K., A survey on: content based image retrieval systems using clustering techniques for large data sets, International Journal of Managing Information Technology 3 (4) (2018) 23.
 12. Abbas J., Qadri S., Idrees M., Awan S., & Khan N. A.99. (2010). Framework for content based image retrieval (Textual based) system, Journal of American Science 6 (9) 704- 707
 13. Kokare, M., Chatterji, B. N., & Biswas, P. K. (2002, May). *A survey on current content based image retrieval methods*. IETE Journal of Research, 48(3–4), 261–271. 10.1080/03772063.2002.11416285
 14. Raghunathan, B., & Acton, S. T., A content based retrieval engine for circuit board inspection, in Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), vol. 1, pp. 104–108 1999. Kobe, Japan.
 15. Jenni, K., Mandala, S., & Sunar, M. S. (2015). *Content based image retrieval using colour strings comparison*. Procedia Computer Science, 50, 374–379. 10.1016/j.procs.2015.04.032
 16. Ali, F., and Hashem, A. (2020, June). *Content Based Image Retrieval (CBIR) by statistical methods*. Baghdad Science Journal, 17 (2(SI)), 694.10.21123/bsj.2020.17.2(SI).0694
 17. Sezavar, A., Farsi, H., & Mohamadzadeh, S. (2019, August). *Content-based image retrieval by combining convolutional neural networks and sparse representation*. Multimedia Tools and Applications, 78(15), 20895–20912. 10.1007/s11042-019- 7321-1
 18. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017, May). *ImageNet classification with deep convolutional neural networks*. Communications of the ACM, 60(6), 84–90. 10.1145/3065386.