BỘ GIÁO DỤC
VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

AN HỒNG SƠN

TRA CỨU ẢNH DỰA VÀO NỘI DUNG VỚI HỌC BIỂU DIỄN VÀ GIẢM CHIỀU DỮ LIỆU

Chuyên ngành: Khoa học máy tính

Mã số: 9 48 01 01

TÓM TẮT LUẬN ÁN TIẾN SĨ NGÀNH KHOA HỌC MÁY TÍNH

Hà Nội - 2023

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ
Viện Hàn lâm Khoa học và Công nghệ Việt Nam

Người hướng dẫn khoa học: PGS.TS. Nguyễn Hữu Quỳnh

Phản biện 1: ……..

Phản biện 2: ……..

Phản biện 3: ……..

Luận án được bảo vệ trước Hội đồng chấm luận án tiến sĩ cấp học viện, họp tại Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi ….. giờ, ngày …... tháng …. năm 2023.

Có thể tìm hiểu luận án tại:

- Thư viện Học viện Khoa học và Công nghệ
- Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của luận án

Trong những năm gần đây, với sự gia tăng nhanh chóng của mạng xã hội cùng với sự phát triển mạnh mẽ của công nghệ 4.0 và các thiết bị di động thông minh, các ứng dụng đa phương tiện đã tạo ra một cơ sở dữ liệu ảnh số khổng lồ. Ảnh số đóng vai trò quan trọng trong nhiều lĩnh vực khác nhau của cuộc sống như viễn thám, thời trang, y học, giáo dục, kiến trúc, phòng chống tội phạm,..... Vì vậy, việc tra cứu nhanh, chính xác một bức ảnh trong một cơ sở dữ liệu ảnh số lớn và đa dạng là một thách thức và nhiệm vụ cấp thiết trong lĩnh vực thị giác máy tính hiện nay.

Trong lĩnh vực thị giác máy tính, Tra cứu ảnh dựa vào nội dung (CBIR-Content-Based Image Retrieval) đang là một trong những hướng được nghiên cứu rất tích cực hiện nay. Mục tiêu của CBIR là tìm kiếm các ảnh dựa trên việc phân tích các nội dung trực quan của ảnh truy vấn [3]. Tuy nhiên, phương pháp này gặp phải vấn đề "khoảng trống ngữ nghĩa" giữa các đặc trưng mức thấp mô tả ảnh và các khái niệm mức cao được con người nhận biết [4], do đó có thể dẫn đến các ảnh không liên quan được trả về. Để khắc phục điều này, nhiều phương pháp đã được đề xuất để chuyển đổi các khái niệm mức cao trong ảnh sang các đặc trưng mức thấp. Các đặc trưng này được phân loại thành các đặc trưng toàn cục (bao gồm màu sắc, hình dạng, kết cấu và thông tin không gian) và các đặc trưng cục bộ tùy thuộc vào phương pháp trích rút đặc trưng [4]. Biểu diễn của các đặc trưng này là nền tảng cho CBIR.

Học máy là một công cụ quan trọng để khai thác các cấu trúc dữ liệu, thu được biểu diễn dữ liệu tốt hơn và khám phá các mẫu dữ liệu ẩn để có thể trích rút được các thông tin liên quan. Trong học máy,

có ba cách tiếp cận chính, bao gồm: học có giám sát, học không giám sát và học bán giám sát. Sự khác nhau của các cách tiếp cận này là ở chỗ sử dụng các mẫu có nhãn trong quá trình học.

Trong những năm gần đây, ở Việt Nam đã có nhiều Nghiên cứu sinh, Nhóm nghiên cứu tiếp cận và khai thác hiệu quả các kỹ thuật học máy cho bài toán CBIR với phản hồi liên quan, giúp thu hẹp "khoảng trống ngữ nghĩa" và cải thiện độ chính xác tra cứu của hệ thống tra cứu ảnh. Tuy nhiên, các công trình này chưa tập trung giải quyết vấn đề cỡ lớp nhỏ và chưa khai thác được thuộc tính thưa dòng của ma trận chiếu. Ngoài ra, tính ưu việt của các kỹ thuật học sâu cho tra cứu ảnh trên tập dữ liệu cỡ lớn, không có nhãn và dữ liệu cao chiều cũng chưa được khai thác. Đây là một định hướng nghiên cứu phù hợp với xu thế nghiên cứu chung của thế giới, mang tính cấp thiết cao và có khả năng ứng dụng hiệu quả trong thực tiễn và đây cũng chính là hướng nghiên cứu mà nghiên cứu sinh đang theo đuổi. Vì vậy Nghiên cứu sinh đã chọn đề tài ***"Tra cứu ảnh dựa vào nội dung với học biểu diễn và giảm chiều dữ liệu"*** làm đề tài luận án của mình.

2. Mục tiêu nghiên cứu của luận án

Nghiên cứu, đề xuất được một số phương pháp cải tiến độ chính xác và thời gian tra cứu đối với những bài toán có cỡ lớp nhỏ, cỡ mẫu nhỏ và dữ liệu chiều cao bằng việc sử dụng kỹ thuật học máy vào quá trình CBIR với phản hồi liên quan.

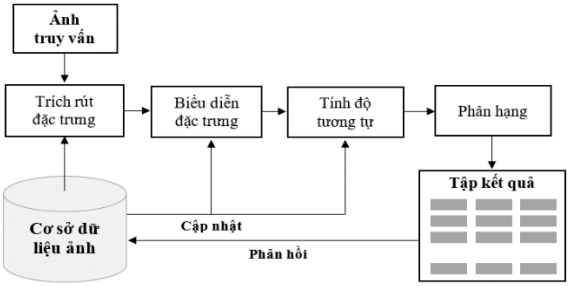
3. Các nội dung nghiên cứu chính của luận án

Luận án tập trung vào nghiên cứu và tìm hiểu một số nội dung chính sau: (***1***) Tra cứu ảnh dựa vào nội dung và biểu diễn các đặc trưng của ảnh; (***2***) Khoảng trống ngữ nghĩa trong CBIR; (***3***) Phản hồi liên quan, kỹ thuật và những thách thức trong phản hồi liên quan; (***4***) Học máy; học sâu, mạng Autoencoder; (***5***) Môi trường thực nghiệm, tập ảnh dữ liệu thực nghiệm và phương pháp đánh giá hiệu năng.

CHƯƠNG 1. TỔNG QUAN VỀ TRA CỨU ẢNH DỰA VÀO NỘI DUNG VỚI PHẢN HỒI LIÊN QUAN

1.1. Tra cứu ảnh dựa vào nội dung

Tra cứu ảnh dựa vào nội dung là một ứng dụng của các kỹ thuật thị giác máy tính đối với bài toán tra cứu ảnh [12].



Hình 1.1. Sơ đồ hệ thống CBIR

Mục tiêu của hệ thống CBIR là sử dụng nội dung trực quan của ảnh để tìm các ảnh quan tâm từ một cơ sở dữ liệu ảnh lớn (nội dung ở đây được hiểu là màu sắc, hình dạng, kết cấu hoặc bất cứ một thông tin nào mà có thể lấy ra từ bản thân ảnh).

1.2. Các đặc trưng mức thấp

Đặc trưng của ảnh có thể được chia thành các đặc trưng toàn cục và đặc trưng cục bộ. Đặc trưng toàn cục, bao gồm: đặc trưng màu, đặc trưng kết cấu, đặc trưng hình và thông tin không gian, trong đó đặc trưng màu được xem là một trong những đặc trưng quan trọng nhất trong tra cứu ảnh. Các đặc trưng cục bộ bao gồm: Biến đổi đặc trưng bất biến tỉ lệ (SIFT), các đặc trưng mạnh và nhanh (SURF), Mẫu nhị phân cục bộ (LBP).

1.3. Lựa chọn đặc trưng

Lựa chọn đặc trưng là quá trình chọn ra tập con các đặc trưng liên quan nhất mà biểu diễn đối tượng dữ liệu một cách hiệu quả nhất. Các đặc trưng này được chọn ra từ các đặc trưng dữ liệu gốc và được

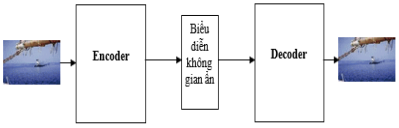
sắp xếp theo thứ tự giảm dần của độ quan trọng. Một số cách tiếp cận đã được đề xuất trong những năm gần đây như: trọng số Fisher [33], nổi trội (Relief) [34], nổi trội F (Relief-F) [35], thông tin tương hỗ (mutual information) [36], điều kiện độc lập của Hilbert Schmidt (HSIC) [37], điểm số Laplace [38]. Trong đó kỹ thuật trọng số Fisher, thuật toán Relief và thuật toán Relief-F được sử dụng phổ biến.

1.4. Trích rút đặc trưng

Việc trích rút đặc trưng là một phương pháp quan trọng để tạo ra các đặc trưng mới dựa trên sự kết hợp hoặc biến đổi nào đó của các đặc trưng gốc. Các phương pháp trích rút đặc trưng cũng giúp thu được các biểu diễn dữ liệu phân biệt hơn. Trích rút đặc trưng được thực hiện thông qua việc chiếu dữ liệu gốc vào các không gian nhúng. Các phương pháp tiêu biểu có thể kể đến bao gồm Phân tích phân biệt tuyến tính (LDA - Linear Discriminant Analysis) [44], Phân tích phân biệt tuyến tính thưa mạnh (RSLDA - Robust Sparse Linear Discriminant Analysis) [41], và trích rút đặc trưng sử dụng giảm gradient (FE_GD - Feature Extraction using Gradient Descent) [43], Phân tích thành phần chính (PCA - Principal Component Analysis) [45].

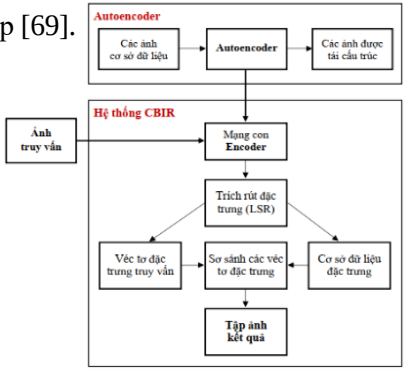
1.5. Học máy cho tra cứu ảnh dựa vào nội dung

Các kỹ thuật học máy thường được áp dụng trong CBIR gồm: (1) Học không giám sát (gồm: Phân cụm K-means và K-means++ [48]); (2) Học có giám sát (gồm: Máy véc tơ hỗ trợ SVM [51] và Mạng nơ ron nhân tạo ANN [55]); (3) Học sâu (gồm: Mạng Autoencoder và Mạng phần dư ResNet [68]); (4) Học kết hợp [69].



Hình 1.2. Mạng Autoencoder

Hình 1.2 mô tả mạng Autoencoder, Hình 1.3 mô tả mô hình tra cứu ảnh dựa vào nội dung với autoencoder.



Hình 1.3. Tích hợp autoencoder với mô hình CBIR

1.6. Cơ chế phản hồi liên quan

Phản hồi liên quan (RF-Relevance Feedback) là một công cụ mạnh được sử dụng phổ biến trong các hệ thống CBIR [76]. Nó được giới thiệu vào đầu những năm 1990, với mục đích đưa người dùng vào quá trình tra cứu để giảm khoảng trống ngữ nghĩa giữa những gì được mô tả bởi các truy vấn (các đặc trưng mức thấp) và những gì người dùng nghĩ. Bằng việc liên tục học thông qua tương tác với người dùng, RF đã cải tiến đáng kể hiệu năng của các hệ thống CBIR [77].

1.7. Đo độ tương tự giữa các ảnh

Đo độ tương tự xác định ảnh nào là ảnh liên quan nhất đến ảnh truy vấn. Do đó, đo độ tương tự ảnh hưởng trực tiếp đến độ chính xác và độ phức tạp tính toán của hệ thống CBIR. Một số độ đo được sử dụng rộng rãi trong CBIR như: Khoảng cách Minkowski; khoảng cách Manhattan; khoảng cách Chessboard; khoảng cách Hamming; khoảng cách lược đồ giao; Khoảng cách Mahalanobis; Khoảng cách Canberra; khoảng cách cosin; thống kê Chi-square; Squared Chord. Chọn độ đo tương tự phù hợp là một nhiệm vụ khó, nhiều nghiên cứu đã thực hiện việc này thông qua các thực nghiệm.

1.8. Một số nghiên cứu về CBIR

1.8.1. Nghiên cứu quốc tế

Năm 2016, Ponomarev và cộng sự trong [90] đã trình bày một hệ thống CBIR dựa trên sự tích hợp của màu sắc, kết cấu và hình dạng. Hạn chế chính của hệ thống là độ phức tạp tính toán tăng lên do tích hợp nhiều đặc trưng. Năm 2017, Srivastava & Khare trong [91] đã phát triển một thuật toán phân tích đa độ phân giải mới giúp phân tích ảnh ở nhiều cấp độ, với các cấp độ khác nắm bắt thông tin mà một cấp độ đã bỏ qua. Cách tiếp cận này dựa trên việc trích rút các đặc trưng kết cấu và hình dạng bằng cách sử dụng bộ mô tả mẫu nhị phân cục bộ (LBP). Một cách tiếp cận CBIR mới được trình bày bằng cách kết hợp các đặc trưng màu, hình dạng và kết cấu do Z.Zhao và cộng sự đề xuất trong [99]. Mặc dù

hệ thống được đề xuất thu được độ chính xác cao, nhưng hiệu năng của hệ thống bị ảnh hưởng khi ảnh truy vấn chứa nhiều đối tượng phức tạp.

Năm 2018, Sajjad và cộng sự trong [92] đã đề xuất một hệ thống CBIR bất biến đối với xoay và thay đổi màu. Hệ thống được đề xuất dựa trên việc ghép các đặc trưng màu và kết cấu để tạo thành một véc tơ đặc trưng chung. Để giảm khoảng trống ngữ nghĩa, Ashraf và cộng sự trong [94] đã đề xuất một hệ thống CBIR kết hợp các đặc trưng màu và cạnh để tạo thành một bộ mô tả đặc trưng. Tuy nhiên, nó vẫn bị thiếu thông tin không gian và không có thông tin về hiệu quả chi phí tính toán. Phadikar và cộng sự trong [100] đã đề xuất một hệ thống CBIR trong miền cosin rời rạc (Discrete Cosine Domain). Mặc dù việc sử dụng thuật toán di truyền có tác động tích cực đến độ chính xác của hệ thống, nhưng nó lại làm tăng thời gian sử dụng.

Năm 2019, Pavithra & Sharmila trong [93] đã đề xuất một phương pháp mới để lựa chọn các điểm hạt giống cho kỹ thuật tra cứu ảnh dựa trên màu trội. Tuy nhiên, phương pháp được đề xuất cần được hợp nhất với các phương pháp trích rút đặc trưng khác (hình dạng, kết cấu và thông tin không gian) để giảm khoảng trống ngữ nghĩa, do cùng một thông tin màu có thể được gán cho các ảnh trong các lớp ngữ nghĩa khác nhau. Một hệ thống CBIR mới được trình bày bởi Bani & Ershad trong [98], dựa trên việc trích rút các đặc trưng kết cấu toàn cục và cục bộ trong cả miền tần số và không gian cũng như các đặc trưng màu trong miền không gian. Hệ thống được đề xuất cho thấy các giá trị có độ chính xác cao và được so sánh với các phương pháp hiện đại khác. Ngoài ra, nó được báo cáo là bất biến với quay và ít nhạy cảm với nhiễu, nhưng nó có thời gian chạy cao do sử dụng các đặc trưng khác nhau.

Năm 2020, Ashraf và cộng sự trong [96] đã phát triển một phương pháp luận cho hệ thống CBIR trên cơ sở kết hợp các đặc trưng mức thấp (kết cấu và màu). Tuy nhiên, lược đồ được đề xuất thiếu thông tin về kết cấu và không gian, như nhiều nghiên cứu khác; Alsmadi và cộng sự trong [97] đã giới thiệu một kỹ thuật tra cứu ảnh dựa trên nội dung mới có lợi thế từ việc kết hợp màu sắc, hình dạng và kết cấu. Kỹ

thuật được đề xuất đã áp dụng thuật toán di truyền, do đó nâng cao chất lượng giải pháp. Tuy nhiên, nó chịu mức độ quan trọng của quá trình và cần lặp lại nhiều lần, làm chậm thời gian tính toán.

1.8.2. Nghiên cứu trong nước

Tại Việt Nam, trong những năm gần đây đã có nhiều công trình nghiên cứu, luận án tiến sĩ liên quan đến bài toán CBIR được công bố, đặc biệt là các công trình nghiên cứu do nhóm nghiên cứu của PGS.TS. Nguyễn Hữu Quỳnh, PGS.TS. Ngô Quốc Tạo, cùng Nghiên cứu sinh và các cộng sự công bố trong các luận án tiến sĩ:

- Năm 2017, Vũ Văn Hiệu đã bảo vệ thành công luận án tiến sĩ "Nghiên cứu một số kỹ thuật phân hạng trong tra cứu ảnh dựa vào nội dung" [101]. Hạn chế là độ chính xác của tập kết quả trong luận án còn thấp do cách tiếp cận của luận án là xét đến một vùng duy nhất chứa các điểm liên quan mà bỏ qua thực tế các ảnh được phân tán trong toàn bộ không gian đặc trưng. Điểm lưu ý ở đây là mặc dù luận án thu các mẫu huấn luyện qua cơ chế phản hồi liên quan nhưng cách tiếp cận của luận án không theo hướng học ma trận chiếu.

- Năm 2019, Đào Thị Thuý Quỳnh đã bảo vệ thành công luận án tiến sĩ "Nâng cao độ chính xác tra cứu ảnh dựa vào nội dung sử dụng kỹ thuật điều chỉnh trọng số hàm khoảng cách" [102]. Hạn chế là phương pháp không xét đến sự không đồng nhất của không gian đặc trưng và không giải quyết vấn đề truy cập xấp xỉ trên không gian non-metric. Điểm lưu ý ở đây là mặc dù luận án thu thập các mẫu huấn luyện qua cơ chế RF nhưng cách tiếp cận của luận án là ma trận chiếu trên cơ sở tận dụng tính địa phương của mỗi vùng điểm đặc trưng.

- Gần đây nhất, năm 2022 NCS. Cù Việt Dũng đã thực hiện luận án tiến sĩ "Nghiên cứu phát triển một số thuật toán tra cứu ảnh dựa vào khái niệm mức cao sử dụng kỹ thuật học sâu" [103]. Mặc dù cách tiếp cận của luận án là học ma trận chiếu với các mẫu huấn luyện được thu từ cơ chế phản hồi liên quan nhưng việc tra cứu ảnh được thực hiện trên không gian chiếu.

Nhìn chung, các công trình này đã tiếp cận và khai thác hiệu quả các kỹ thuật học máy cho CBIR và thực nghiệm trên các tập dữ liệu ảnh chuyên nghiệp, phổ biến. Tuy nhiên, các công trình này chưa khai thác được thuộc tính thưa dòng của ma trận chiếu và học biểu diễn ảnh theo tiếp cận học sâu. Đây là một hướng nghiên cứu thiết thực, có tính khả thi cao mà Nghiên cứu sinh hướng đến trong các nội dung nghiên cứu tại luận án này.

1.9. Tổ chức thực nghiệm và đánh giá hiệu năng

1.9.1. Cơ sở dữ liệu ảnh thực nghiệm

Dữ liệu thực nghiệm được sử dụng trong luận án này là các tập CSDL ảnh chuyên nghiệp, đã được sử dụng rộng rãi để đánh giá hiệu năng của hệ thống CBIR [104], bao gồm tập CSDL ảnh COREL (Hình 1.7), CIFAR-100 (Hình 1.8).

Hình 1.7. Một số ảnh đại diện của 5 khái niệm ngữ nghĩa trong Corel

Hình 1.8. Một số ảnh đại diện của 5 khái niệm ngữ nghĩa trong CIFAR-100

1.9.2. Phương pháp đánh giá hiệu năng

Trong luận án này, thước đo được sử dụng để đánh giá hiệu năng của các phương pháp đề xuất là: *AP* và *mAP*.

$$\mathrm{AP} = \frac{\sum_{k=1}^{N} P(k).rel(k)}{R} \qquad \textbf{(1.25)}$$

$$mAP = \frac{1}{Q}\sum_{q=1}^{Q} AP(q) \qquad \textbf{(1.26)}$$

1.10. Kết luận Chương 1

Trong chương này, luận án đã hệ thống lại những kiến thức lý thuyết cơ sở và nghiên cứu liên quan đến CBIR, đồng thời phân tích nghiên cứu liên quan đến các giai đoạn trong CBIR để thấy được ưu điểm và hạn chế của các nghiên cứu hiện nay, làm cơ sở khẳng định tính khả thi của hướng nghiên cứu và xác định các nội dung cần giải quyết ở các chương tiếp theo của luận án.

CHƯƠNG 2. PHƯƠNG PHÁP TRA CỨU ẢNH VỚI PHÂN TÍCH PHÂN BIỆT THƯA

2.1. Giới thiệu

Bài toán tra cứu ảnh với phản hồi liên quan sử dụng cách tiếp cận phân lớp chỉ bao gồm hai lớp là âm và dương, do đó nó có một số vấn đề sau: (1) Số các mẫu thường quá nhỏ so với chiều của không gian đặc trưng [115], (2) Số các mẫu âm thường nhiều hơn số các mẫu dương rất nhiều [115], và (3) Số các lớp là quá nhỏ, dẫn đến số các hướng chiếu phải nhỏ bởi vì số các hướng chiếu có liên quan chặt chẽ đến số các lớp. Để giải quyết 3 vấn đề này, luận án đề xuất một phương pháp tra cứu ảnh có giám sát mới, kết hợp mô hình trích rút đặc trưng quan trọng dựa trên phương pháp RSLDA với mô hình phân lớp trong hệ thống CBIR nhằm cải tiến độ chính xác và thời gian truy vấn. Phương pháp có tên **SDAIR** (Sparse Discriminant Analysis for Image Retrieval).

SDAIR có các đặc tính sau: (1) Mô hình rất mềm dẻo, có thể áp dụng với bất kỳ độ đo tương tự ảnh nào, mô hình học lựa chọn đặc trưng nào, và mô hình học phân lớp nào; (2) Không bị ảnh hưởng bởi vấn đề cỡ lớp nhỏ, trong khi nó vẫn loại đi được các đặc trưng dư thừa và không liên quan, và tận dụng được thông tin phân biệt; (3) Không đòi hỏi số các mẫu dương phải đủ lớn bởi vì nó có thể cung cấp cơ chế tự động bổ sung mẫu dương vào tập mẫu huấn luyện (không đòi hỏi phải huấn luyện lại mô hình học chiếu); (4) Hỗ trợ đồng thời cho hai nhiệm vụ đó là lựa chọn tập đặc trưng quan trọng và bổ sung mẫu huấn luyện dương.

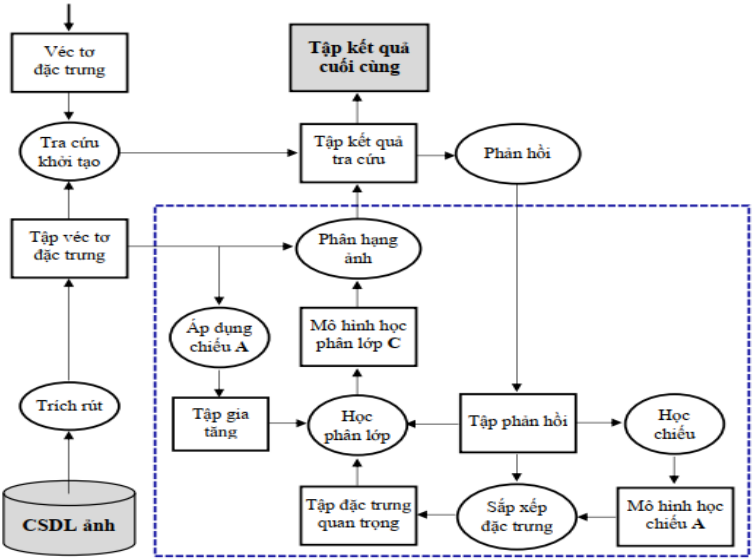
2.2. Phương pháp tra cứu ảnh được đề xuất

2.2.1. Mô hình của phương pháp

Mô hình tra cứu ảnh được đề xuất trong Hình 2.1. Quá trình tra cứu bắt đầu bằng việc trích rút đặc trưng của ảnh truy vấn. Sử dụng các véc tơ đặc trưng này cùng với một độ đo tương tự được xác định trước để đo độ tương tự giữa ảnh truy vấn và các ảnh CSDL. Sau đó,

một tập các ảnh liên quan với ảnh truy vấn được lựa chọn và tập này được sắp xếp theo thứ tự giảm dần của độ tương tự để thu được tập kết quả tra cứu.

Người dùng phản hồi trên tập kết quả tra cứu để nhận được tập phản hồi, cũng là tập mẫu huấn luyện. Trên cơ sở tập mẫu huấn luyện này, thuật toán học chiếu được thực hiện để thu được mô hình học chiếu A. Sử dụng mô hình học chiếu A trên tập phản hồi và sắp xếp các đặc trưng theo thứ tự giảm dần của độ quan trọng để thu được tập đặc trưng quan trọng. Để giải quyết vấn đề cỡ mẫu nhỏ và số mẫu dương ít hơn số mẫu âm trong bài toán CBIR với phản hồi liên quan, mô hình tự động bổ sung mẫu dương thông qua áp dụng chiếu A vừa được học lên tập đặc trưng để thu được tập gia tăng. Thu các đặc trưng quan trọng trên cả hai tập phản hồi và gia tăng để tạo ra tập huấn luyện cho học phân lớp, do đó thu được mô hình học phân lớp C. Phân hạng các ảnh sẽ được thực hiện theo mô hình học phân lớp C để được tập kết quả tra cứu. Quá trình này sẽ được lặp lại nếu người dùng chưa thỏa mãn với kết quả tra cứu, ngược lại thu được tập kết quả cuối cùng.



Hình 2.1. Mô hình của phương pháp tra cứu ảnh được đề xuất

2.2.2. Mô hình học chiếu cho lựa chọn tập đặc trưng quan trọng

RSLDA [41] là một phương pháp trích rút đặc trưng dựa vào LDA. Nó cực tiểu $\ell_{2,1}$ norm của ma trận chiếu tuyến tính Q. RSLDA có thể khôi phục dữ liệu ban đầu từ dữ liệu được chiếu chiều thấp.

Nhằm trích rút các đặc trưng mà vẫn bảo toàn được năng lượng chính của dữ liệu, RSLDA giải bài toán tối ưu sau:

$$\min_{P,Q,E} Tr(Q^T(S_w - \lambda S_b)Q) + \lambda_1\|Q\|_{2,1} + \lambda_2\|E\|_1 \qquad \textbf{(2.6)}$$

$$\text{Thoả mãn} \quad X = PQ^TX + E,\ P^TP = I$$

Lấy động lực để khắc phục hạn chế của LDA, và kế thừa các ưu điểm của phương pháp RSLDA, luận án đề xuất một mô hình học bằng việc bổ sung một số hạng để khớp các nhãn lớp (các mẫu có cùng nhãn trong không gian chiếu sẽ gần nhau hơn trong khi các mẫu có nhãn khác nhau sẽ cách xa nhau hơn) giúp tăng tính phân lớp của ma trận chiếu thu được. Cực tiểu hàm mục tiêu ở (2.7) dưới đây.

$$\min_{P,A,E} Tr(A^T(S_w - \lambda S_b)A) + \lambda_1\|A\|_{2,1} + \lambda_2\|E\|_1 + \frac{1}{2}\|Y - AX\|_F^2 \quad \textbf{(2.7)}$$

$$\text{Thoả mãn} \quad X = PA^TX + E,\ P^TP = I$$

Thuật toán 2.1: Chọn tập đặc trưng quan trọng

Input: - Ma trận mẫu huấn luyện X, ma trận nhãn Y
- Các tham số λ_1, λ_2, số đặc trưng quan trọng k

Output: - Ma trận chiếu A
- Ma trận đặc trưng quan trọng X_k

Step 1: Tính S_b theo công thức (2.2); Tính và S_w theo công thức (2.3)

Step 2: Giải bài toán tối ưu (2.7) theo [132] để có ma trận chiếu A

Step 3: Tính $\|a_i\|_2, i = 1,2,\dots,m$ của A

Step 4: Sắp xếp m dòng của X theo thứ tự giảm dần của $\|a_i\|_2$. Xây dựng X_k gồm k dòng trên đỉnh của X.

Step 5: Return A và X_k

2.2.3. Mô hình học cho phân lớp

Phần này kế thừa giải pháp xử lý của vấn đề cỡ mẫu nhỏ trong Thuật toán 2.1 và tập trung vào giải quyết pha phân lớp của bài toán tra cứu ảnh với phản hồi liên quan.

Để giải quyết được bài toán cỡ lớp nhỏ ở trên, luận án đề xuất mô hình học phân lớp nhưng nó được thực hiện trên không gian đặc trưng gốc. Khi thực hiện phân lớp trên không gian đặc trưng gốc, phải đối mặt với vấn đề về số chiều của không gian đặc trưng cao, do đó ta loại đi các đặc trưng dư thừa (xem Thuật toán 2.1).

Thuật toán phân lớp được tóm tắt trong Thuật toán 2.2 sau:

Thuật toán 2.2: Xây dựng mô hình phân lớp

Input: - Ma trận mẫu huấn luyện X, ma trận nhãn L
- Mô hình học chiếu A ;
- Tập véc tơ đặc trưng F
- Ma trận đặc trưng quan trọng X_k

Output: Mô hình học phân lớp R

Step 1: Áp dụng mô hình học chiếu A lên tập véc tơ đặc trưng F.

Step 2: Xây dựng ma trận gia tăng $X^{(e)}$ bao gồm e điểm x_i tương ứng với e điểm y_i mà là lân cận của $y_i^{(q)}$. Xây dựng ma trận nhãn $L^{(e)}$ bao gồm e nhãn dương của $x_i \in X^{(e)}$.

Step 3: Gộp ma trận $X^{(e)}$ vào ma trận X theo nguyên tắc cột đầu tiên của $X^{(e)}$ xếp ở bên phải cột cuối cùng của X. Tương tự trong việc gộp ma trận $L^{(e)}$ vào L.

Step 4: Huấn luyện phương pháp học phân lớp trên X và L.

Step 5: Return mô hình học phân lớp R.

2.2.4. Thuật toán tra cứu ảnh đề xuất

Thuật toán được đề xuất gọi Thuật toán 2.1 trong bước 2 (Step 2.2) để giảm chiều và thu được tập đặc trưng quan trọng. Bước này giúp giải quyết vấn đề chiều cao và hỗ trợ giải quyết vấn đề cỡ lớp nhỏ

(trong Thuật toán 2.2) của bài toán tra cứu ảnh với phản hồi liên quan, mà sử dụng phân lớp. Bước 3 (Step 2.3) giải quyết vấn đề cỡ lớp nhỏ, cỡ mẫu nhỏ và bị lệch thông qua việc gọi Thuật toán 2.2.

Thuật toán đề xuất được tóm tắt trong Thuật toán 2.3 như sau:

Thuật toán 2.3: SDAIR

Input: **F**: tập đặc trưng của các ảnh cơ sở dữ liệu, **q**: véc tơ ảnh truy vấn, ***N***: số các ảnh tại mỗi vòng lặp.

Output: **S**: tập kết quả.

Step 1: Tra cứu ảnh với q để được tập kết quả khởi tạo và lấy *N* véc tơ ảnh ở top để được tập kết quả *I*

Step 2:

Repeat

Step 2.1: Người dùng phản hồi trên tập *I* để có tập phản hồi RF

Step 2.2: Thực hiện Thuật toán 2.1 để có ma trận đặc trưng qt X_k

Step 2.3: Thực hiện Thuật toán 2.2 để có mô hình học phân lớp C

Step 2.4: Phân hạng tập đặc trưng F theo mô hình học phân lớp C để được danh sách phân hạng

Step 2.5: Lấy *N* ảnh ở trên đỉnh của danh sách phân hạng trong Step 2.4 làm tập ảnh kết quả S

Until (User stops responding)

Step 3: Return S.

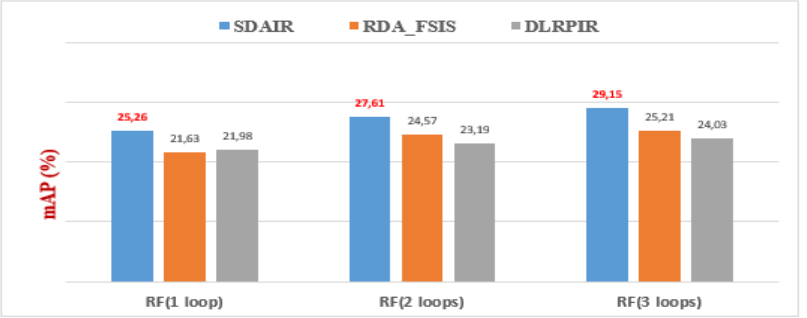
2.3. Kết quả thực nghiệm

Thực nghiệm đầu tiên là so sánh phương pháp đề xuất với các phương pháp tra cứu ảnh tiêu biểu, để chỉ ra rằng phương pháp đề xuất có độ chính xác tổng thể cao hơn các phương pháp còn lại. Thực nghiệm thứ hai là để kiểm tra hiệu quả của việc loại bỏ các đặc trưng dưa thừa và không liên quan, đồng thời giải quyết vấn đề cỡ lớp nhỏ trên cơ sở dữ liệu CIFAR-100. Độ đo mAP (trong 1.9.3) cũng được sử dụng để đánh giá độ chính xác của phương pháp đề xuất.

Phương pháp DLRPIR và RDA_FSIS [42] được sử dụng để so sánh với phương pháp đề xuất là vì nó sử dụng cùng một độ đo tương tự và cơ chế phản hồi như phương pháp đề xuất, đồng thời nó sử dụng phương pháp chiếu hạng thấp phân biệt để chiếu dữ liệu gốc sang một không gian chiếu, sau đó thực hiện phân lớp trên không gian chiếu này để phân hạng các ảnh.

2.3.1. Kiểm tra hiệu năng toàn bộ của phương pháp đề xuất

Hình 2.8 chỉ ra độ chính xác trung bình của ba phương pháp tại top 100 ảnh cho ba lần lặp đầu tiên. Với các kết quả này, cho thấy rằng, độ chính xác của phương pháp RDA_FSIS cao hơn DLRPIR là bởi vì nó học được một ma trận chiếu phân biệt thưa theo cấu trúc của từng lớp và giảm vấn đề cỡ lớp nhỏ. Độ chính xác của phương pháp đề xuất là cao nhất trong ba phương pháp bởi vì nó loại đi các đặc trưng dư thừa và không liên quan. Bên cạnh đó, nó cũng giải quyết hiệu quả vấn đề cỡ lớp nhỏ.



Hình 2.8. mAP của ba phương pháp trên top 100

2.3.2. Thực nghiệm về hiệu quả tra cứu ảnh khi loại bỏ các đặc trưng dư thừa và giải quyết vấn đề cỡ lớp nhỏ

Luận án thiết kế ba kịch bản thực nghiệm như sau:

Kịch bản (**1**): So sánh hiệu quả tra cứu mà không sử dụng phản hồi (chỉ sử dụng Euclide) trên không gian gồm 1,305 chiều và trên không gian gốc nhưng loại đi các chiều dư thừa và không quan trọng.

Kịch bản (**2**): So sánh hiệu quả tra cứu mà không sử dụng phản hồi (chỉ sử dụng Euclide) trên không gian gốc (nhưng loại đi các chiều dư thừa và không quan trọng) và trên không gian chiếu.

Kịch bản (**3**): So sánh hiệu quả tra cứu sử dụng phản hồi trên các không gian bao gồm: *(1)* không gian gốc ban đầu (có 1,305 chiều); *(2)* không gian gốc (nhưng loại đi các chiều dư thừa và không quan trọng); và *(3)* không gian chiếu. Trong kịch bản này, mô hình SVM được sử dụng để phân hạng các ảnh và thu về tập kết quả tra cứu.

Số chiều mà luận án thực nghiệm trong cả ba kịch bản ở trên bao gồm: 30 chiều gốc (loại đi 1,275 chiều gốc), 20 chiều gốc (loại đi 1,285 chiều gốc), và 10 chiều gốc (loại đi 1,295 chiều gốc). Bảng 2.2, 2.3 và 2.4 là kết quả tương ứng với các kịch bản (1), (2), và (3).

Bảng 2.2. Kết quả tra cứu ảnh theo kịch bản (1)

Phương pháp	**OIR_i**				
Số chiều	1305	128	30	20	10
mAP (%)	16,07	**18,27**	16,63	16,15	15,6

Bảng 2.3. Kết quả tra cứu ảnh theo kịch bản (2)

Phương pháp	**OIR_i**				**PIR_i**			
Số chiều	128	30	20	10	128	30	20	10
mAP (%)	**18,27**	**16,63**	**16,15**	**15,6**	17,21	15,68	15,3	15,05

Bảng 2.4. Kết quả tra cứu ảnh theo kịch bản (3)

Phương pháp	**OIRRF_i**					**PIRRF_i**			
Số chiều	1305	128	30	20	10	128	30	20	10
mAP (%)	20,3	**25,26**	20,56	19,16	18,63	20,9	19,76	18,96	18,63

Nhìn vào Bảng 2.2 thấy rằng, độ chính xác khi lựa chọn 128 chiều là cao nhất trong số các chiều gồm 128, 30, 20, và 10. Điều này là minh chứng để khẳng định hiệu quả khi loại bỏ các đặc trưng dư thừa và không liên quan của phương pháp đề xuất.

Bảng 2.3, độ chính xác của phương pháp đề xuất trên không gian gốc là cao hơn độ chính xác trên không gian chiếu ở tất cả các chiều bao gồm 128, 30, 20, và 10. Lý do của việc này là bởi vì trên không gian gốc, có thể xác định được đặc trưng nào là quan trọng nhất để giữ lại trong khi trên không gian chiếu, không biết được đặc trưng nào là quan trọng để giữ lại, dẫn đến có thể giữ lại những đặc trưng ít quan trọng và loại đi những đặc trưng quan trọng.

Số liệu trên Bảng 2.4 cho thấy rằng, ở các chiều 128, 30, 20, và 10, độ chính xác của phương pháp đề xuất trên không gian gốc luôn cao hơn trên không gian chiếu. Lý do của điều này là ngoài việc loại đi được các đặc trưng dưa thừa và không liên quan, nó còn giảm được sự ảnh hưởng của vấn đề cỡ lớp nhỏ.

Bảng 2.5 ở dưới chỉ ra thời gian truy vấn của phương pháp tra cứu ảnh trên không gian gốc và không gian chiếu.

Bảng 2.5. Thời gian truy vấn theo số chiều trên không gian gốc và không gian chiếu

Phương pháp	**Thời gian chạy của OIR_i**					**Thời gian chạy của PIR_i**			
Số chiều	1305	128	30	20	10	128	30	20	10
Thời gian (s)	0.5531	**0.35**	0.20	0.19	0.18	0.44	0.49	0.42	0.34

2.4. Kết luận Chương 2

Chương này, luận án đã đề xuất được một mô hình mềm dẻo, bằng cơ chế học tự động bổ sung mẫu dương vào tập huấn luyện, không đòi hỏi số các mẫu dương phải đủ lớn, ngoài ra nó có thể phục vụ đồng thời cho hai nhiệm vụ đó là lựa chọn tập đặc trưng quan trọng và bổ sung mẫu huấn luyện dương. Các kết quả thực nghiệm trên cơ sở dữ liệu CIFAR-100 đã cho thấy rằng phương pháp đề xuất có thể cải tiến hiệu năng cho bài toán tra cứu ảnh với phản hồi liên quan, nơi mà cỡ mẫu nhỏ, cỡ lớp nhỏ, và dữ liệu có chiều cao.

Các đóng góp chính của chương này đã được công bố trong công trình [**CT4, CT2**].

CHƯƠNG 3. HỌC CÁC BIỂU DIỄN ẢNH VỚI MẠNG NƠ RON TÍCH CHẬP SÂU AUTOENCODER CHO TRA CỨU ẢNH VỚI PHẢN HỒI LIÊN QUAN

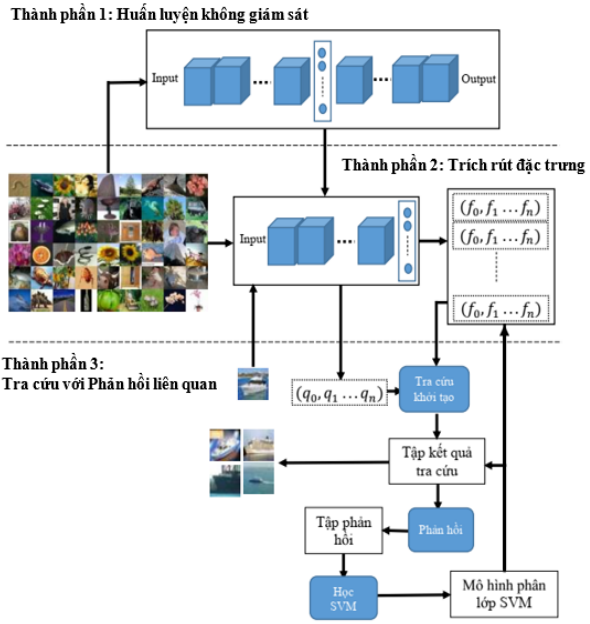
3.1. Giới thiệu

Hiệu năng của bất cứ phương pháp CBIR nào cũng phụ thuộc chính vào biểu diễn mô tả đặc trưng của ảnh và cũng đều được kỳ vọng là có khả năng phân biệt, mạnh và chiều thấp. Đặc trưng được thiết kế thủ công cho tra cứu ảnh là một lĩnh vực nghiên cứu rất tích cực, tuy nhiên hiệu năng của nó bị giới hạn do thiết kế thủ công không thể biểu diễn các đặc tính của ảnh theo một cách chính xác [135].

Để giải quyết các hạn chế được nêu ở trên, luận án đề xuất một phương pháp bán giám sát có tên là AIR dựa trên ba thành phần (mạng nơ ron tích chập autoencoder, trích rút đặc trưng ảnh và phân lớp SVM trong phản hồi liên quan). Phương pháp AIR khắc phục được hai vấn đề: (1) khả năng phân biệt các đặc trưng kém của các phương pháp trước do được tích hợp cơ chế RF và phân hạng qua máy véc tơ hỗ trợ SVM và (2) giảm nhẹ vấn đề vanishing/exploding gradients và độ phức tạp tính toán thông qua việc sử dụng các kết nối tắt (shortcut connections) trong kiến trúc autoencoder và dẫn đến có thể sử dụng các autoencoder sâu.

3.2. Phương pháp đề xuất

Phương pháp đề xuất gồm ba thành phần. Thành phần thứ nhất là huấn luyện không giám sát mạng nơ ron autoencoder sâu trên một tập con của tập ảnh. Thành phần thứ hai là áp dụng mô hình học từ thành phần thứ nhất để trích rút các đặc trưng thấp chiều từ tập ảnh CSDL *(ở đây, cả thành phần thứ nhất và thứ hai được thực hiện offline)*. Thành phần thứ ba là tra cứu các ảnh tương tự với ảnh truy vấn dựa vào phản hồi liên quan. Mô hình autoencoder được huấn luyện trên một tập con của tập CSDL ảnh CIFAR-100.



Hình 3.1. Mô hình của phương pháp tra cứu ảnh đề xuất

3.2.1. Học các biểu diễn ảnh với mạng nơ ron tích chập autoencoder

3.2.1.1. Mạng nơ ron tích chập sâu autoencoder

Đầu tiên, ảnh đầu vào được mã hóa mà mỗi thời điểm một mảng vá $d \times d$ pixel $p_i, i = 1,2, \dots k$, được lựa chọn ra từ ảnh đầu vào, và sau đó trọng số w_j của nhân chập j được sử dụng cho tính toán tích chập. Cuối cùng giá trị nơ ron $a_{ij}, j = 1,2, \dots, m$ được tính toán từ lớp đầu ra.

$$a_{ij} = f(p_i) = \sigma(w_j . p_i + b) \qquad \textbf{(3.1)}$$

$$RElu(p) = \begin{cases} p \text{ nếu } p \geq 0 \\ 0 \text{ nếu } p < 0 \end{cases} \qquad \textbf{(3.2)}$$

Sau đó, đầu ra o_{ij} từ bộ giải mã tích chập được mã hóa mà p_i được tái cấu trúc qua a_{ij} để tạo ra $\hat{p}_i$.

$$\hat{p}_i = f'(a_{ij}) = \emptyset(w_i . a_{ij} + \hat{b}) \qquad \textbf{(3.3)}$$

$\hat{p}_i$ được tạo ra sau mỗi mã hóa và giải mã tích chập. Ta nhận được mảng vá P mà thu được từ toán tử tái cấu trúc. Sử dụng sai số bình phương trung bình giữa mảng vá gốc của ảnh đầu vào $p_i, i = 1,2, \dots k$

và mảng vá tái cấu trúc của ảnh $\hat{p}_i, i = 1,2,..,k$.

Hàm chi phí được mô tả trong phương trình (3.4), và sai số tái cấu trúc được mô tả trong phương trình (3.5).

$$L(\theta) = \frac{1}{k}\sum_{i=1}^{k} E(p_i, \hat{p}_i) \quad \textbf{(3.4)}$$

$$E(p_i, \hat{p}_i) = \|p_i - \hat{p}_i\|^2 = \|p_i - \emptyset(\sigma(p_i))\|^2 \quad \textbf{(3.5)}$$

3.2.1.2. Lớp pooling

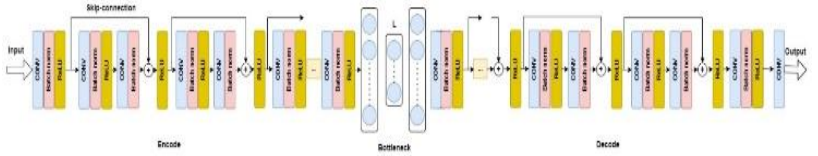
Giống như trong CNN, lớp tích chập được kết nối với lớp pooling [92]. Trong kiến trúc mạng nơ ron tích chập, lớp gộp cực đại (max pooling) được đặt sau lớp tích chập autoencoder, cụ thể:

$$a_j^i = \max(p_j^i) \quad \textbf{(3.6)}$$

Trong phương trình (3.6), p_j^i biểu diễn vùng thứ i của bản đồ đặc trưng thứ j, và a_j^i biểu diễn nơ ron thứ i của bản đồ đặc trưng thứ j.

3.2.1.3. Kiến trúc mạng tích chập autoencoder

Các mạng nơ ron sâu gặp phải vấn đề vanishing/exploding gradients và độ phức tạp tính toán. Bởi vì các autoencoder có nhiều lớp convolutional and deconvolutional nên bị mất mát thông tin và làm giảm hiệu năng khi tái cấu trúc các ảnh. Lấy cảm hứng từ mạng phần dư bao gồm các shortcut connections [75], ta bổ sung shortcut connections vào mạng autoencoder như trên Hình 3.2. Các kết nối này giúp gửi trực tiếp các bản đồ đặc trưng từ lớp đầu tiên của encoder đến một số lớp sau.



Hình 3.2. Kiến trúc mạng autoencoder đề xuất cho trích rút đặc trưng

3.2.2. Tra cứu ảnh với phản hồi liên quan dựa vào máy véc tơ hỗ trợ

3.2.2.1. Máy véc tơ hỗ trợ (SVM)

Trong phần này, luận án chọn máy véc tơ hỗ trợ SVM [39] cho việc phân lớp và phân hạng các ảnh là bởi vì: *Thứ nhất*, nó là một bộ phân lớp mạnh, đặc biệt cho phân lớp nhị phân, mà bài toán tra cứu

ảnh với phản hồi liên quan là bài toán có hai lớp. *Thứ hai*, thông qua siêu phẳng tối ưu tìm được, có thể sử dụng khoảng cách từ mỗi mẫu đến siêu phẳng tối ưu làm giá trị để phân hạng các ảnh.

3.2.2.2. Tra cứu ảnh

Như mô hình phương pháp trên Hình 3.1, sau khi huấn luyện được mô hình mạng nơ ron tích chập autoencoder ở Thành phần 1, ta tiến hành bỏ phần decoder đi và giữ lại phần encoder để có mô hình học như trong Thành phần 2. Sử dụng mô hình học trong Thành phần 2 của mô hình cho trích rút các véc tơ đặc trưng thấp chiều để thu được tập gồm n véc tơ đặc trưng $(f_0, f_1 \ ... \ f_n)$.

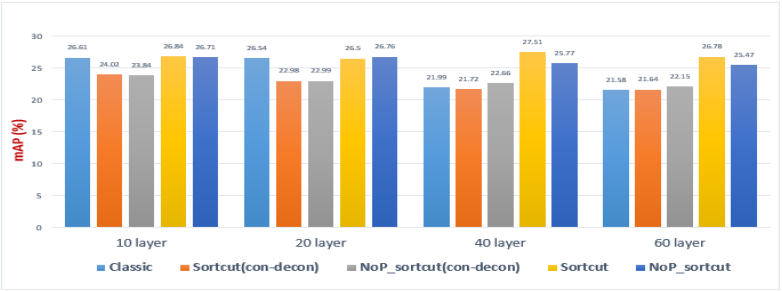
Trong quá trình tra cứu như trong Thành phần 3 của mô hình, người dùng cung cấp một ảnh truy vấn q, véc tơ của ảnh truy vấn sẽ được đưa qua mô hình học encoder để có véc tơ đặc trưng của ảnh truy vấn $(q_0, q_1 \ \ldots . q_n)$. Quá trình tra cứu khởi tạo sẽ so sánh (dùng Euclide) véc tơ của ảnh truy vấn với véc tơ của mỗi ảnh CSDL để thu được tập kết quả tra cứu. Trên tập kết quả này, người dùng phản hồi để thu được tập phản hồi (tập phản hồi này bao gồm các mẫu có nhãn âm và dương, nó cũng là tập huấn luyện). Học SVM được áp dụng trên tập huấn luyện để thu được mô hình phân lớp SVM. Áp dụng mô hình phân lớp trên tập véc tơ đặc trưng CSDL ảnh: những ảnh được dự đoán có nhãn dương mà có khoảng cách xa nhất từ siêu phẳng tối ưu) sẽ được xếp ở vị trí số một của danh sách kết quả, những ảnh được dự đoán có nhãn dương mà có khoảng cách xa thứ nhì từ siêu phẳng tối ưu sẽ được xếp ở vị trí số hai của danh sách kết quả, ... Quá trình này lặp đi lặp lại cho đến khi người dùng dừng phản hồi.

3.3. Đánh giá thực nghiệm

3.3.1. Các kết quả trên tập dữ liệu ảnh CIFAR-100

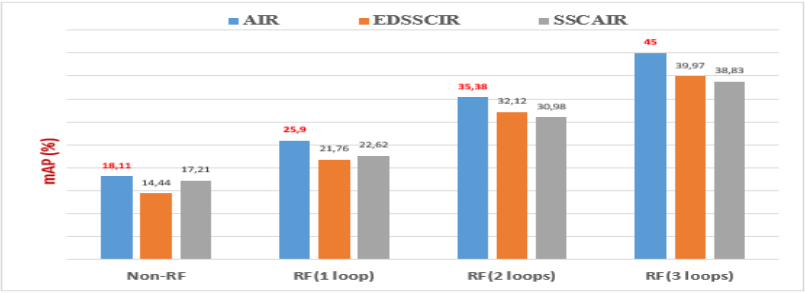
Hình 3.9 cho thấy số lớp tối ưu của kiến trúc mạng autoencoder cho tra cứu ảnh trên tập CIFAR-100 là 40 lớp và cấu hình mạng sử dụng lớp pooling có hiệu quả cho kiến trúc mạng càng sâu. Trong số 5 cấu hình, hai cấu hình trong kiến trúc mạng được đề xuất cho kết quả cao

nhất trên toàn bộ 20, 40, và 60 lớp. Điều này chứng tỏ rằng việc sử dụng shortcut connections không đối xứng vào autoencoder để tạo ra các mạng sâu autoencoder là hiệu quả trên tập CIFAR-100.



Hình 3.9. Kết quả tra cứu ảnh theo các độ sâu khác nhau của mạng autoencoder trên tập CIFAR-100

Hình 3.10 chỉ ra mAP của bốn phương pháp gồm Baseline (Non-RF), AIR, EDSSCIR, và SSCAIR cho ba lần lặp phản hồi đầu tiên. Trong đó, phương pháp Baseline cho độ chính xác thấp nhất. Lý do của việc này là do phương pháp Baseline không có cơ chế học, nó chỉ tính khoảng cách Euclide giữa véc tơ đặc trưng của ảnh truy vấn và ảnh CSDL. Phương pháp AIR thực hiện tốt hơn hai phương pháp còn lại trên tất cả các lần lặp. Hiệu năng của AIR là tốt hơn đáng kể so với Baseline, nó chỉ ra rằng các phản hồi liên quan được người dùng cung cấp là rất hữu ích trong cải tiến hiệu năng tra cứu. AIR thực hiện tốt hơn EDSSCIR là bởi vì AIR thu được biểu diễn đặc trưng tốt.



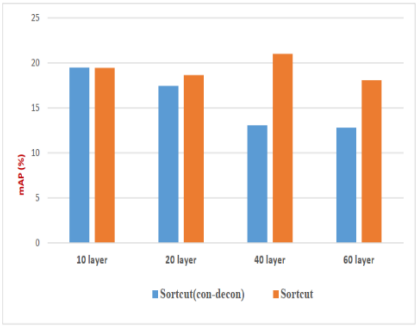
Hình 3.10. So sánh hiệu năng (mAP) cho ba lần lặp đầu tiên

Bảng 3.4. Thời gian thực hiện truy vấn của AIR trên CIFAR-100

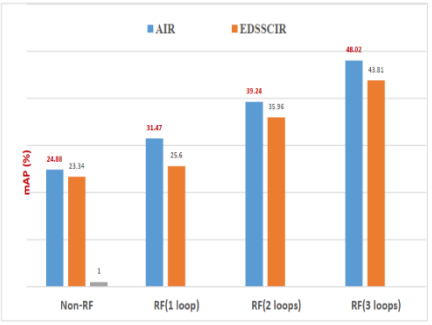
Vòng lặp phản hồi	**Thời gian trung bình cho 1 truy vấn của AIR**		
	Shortcut(con-decon) (s)	**Shortcut (s)**	**Classic (s)**
Không có phản hồi	0.2449	0.2650	0.2335
Vòng lặp thứ nhất	25.5623	28.1375	24.0926
Vòng lặp thứ hai	26.2186	28.9882	24.4392
Vòng lặp thứ ba	27.2913	29.1830	24.5538

Bảng 3.4 cho thấy phương pháp sử dụng Shortcut có thời gian cao hơn Shortcut(con-decon) (cao hơn ~2s). Nguyên nhân của việc này là bởi vì nó phải dành thời gian để tính thêm kết nối tắt.

3.3.2. Các kết quả trên tập dữ liệu ảnh Corel



Hình 3.11. Kết quả tra cứu ảnh theo các độ sâu khác nhau của mạng autoencoder trên tập COREL



Hình 3.12. So sánh hiệu năng (dưới dạng mAP) của ba phương pháp cho ba lần lặp đầu tiên

Hình 3.11, cho thấy kiến trúc mạng được đề xuất cho kết quả cao nhất trên 40, và 60 lớp, riêng 20 lớp là cho kết quả xấp xỉ nhau. Điều này chứng tỏ rằng việc sử dụng shortcut connections không đối xứng vào autoencoder để tạo ra các mạng sâu autoencoder cho đối sánh ảnh là hiệu quả trên tập COREL.

Hình 3.12 chỉ ra mAP của ba phương pháp gồm Baseline (Non-RF), AIR, EDSSCIR cho ba lần lặp phản hồi đầu tiên. Từ Hình 3.9 thấy rằng, phương pháp Baseline cho độ chính xác thấp nhất. Lý

do của việc này là do phương pháp Baseline không có cơ chế học, nó chỉ tính khoảng cách Euclide giữa véc tơ đặc trưng của ảnh truy vấn và ảnh cơ sở dữ liệu. Phương pháp được đề xuất AIR thực hiện tốt hơn hai phương pháp còn lại trên tất cả các lần lặp. Hiệu năng của phương pháp AIR là tốt hơn đáng kể so với Baseline, nó chỉ ra rằng các phản hồi liên quan được người dùng cung cấp là rất hữu ích trong cải tiến hiệu năng tra cứu. AIR thực hiện tốt hơn EDSSCIR là bởi vì AIR thu được biểu diễn đặc trưng tốt.

Bảng 3.5 cho thấy phương pháp sử dụng Shortcut có thời gian cao hơn Shortcut(con-decon) (cao hơn ~0.02s). Nguyên nhân của việc cao hơn là bởi vì nó phải dành thời gian để tính thêm kết nối tắt.

Bảng 3.5. Thời gian thực hiện truy vấn của AIR trên COREL

Vòng lặp phản hồi	**Thời gian TB cho 1 truy vấn của AIR với cấu hình**		
	Shortcut(con-decon) (s)	**Shortcut (s)**	**Classic (s)**
Không có phản hồi	0.1289	0.1468	0.0457
Vòng lặp thứ nhất	5.5781	5.5734	4.8175
Vòng lặp thứ hai	5.6410	5.6508	4.8858
Vòng lặp thứ ba	5.8743	5.8919	4.8108

3.4. Kết luận Chương 3

Trong chương này, luận án trình bày một phương pháp hiệu quả cho tra cứu ảnh. Phương pháp này đã khắc phục được hai vấn đề: *thứ nhất*, khả năng phân biệt hạn chế của các phương pháp đã có và *thứ hai*, giảm nhẹ vấn đề vanishing/exploding gradients và quá trình hội tụ nhanh. Mô hình mạng nơ ron tích chập sâu autoencoder được tận dụng để học các biểu diễn đặc trưng hiệu quả cho tra cứu ảnh thông qua việc sử dụng shortcut connections trong kiến trúc autoencoder.

Các đóng góp chính của chương này đã được Nghiên cứu sinh công bố trong các công trình [**CT1, CT3**].

KẾT LUẬN

Luận án đã xác định được hướng nghiên cứu cần tập trung đó là: tiếp cận sử dụng học máy (đặc biệt là học sâu) vào quá trình tra cứu ảnh với phản hồi liên quan để rút ngắn khoảng trống ngữ nghĩa, nâng cao độ chính xác và tốc độ tra cứu trong CBIR đối với các bài toán có cỡ lớp nhỏ, cỡ mẫu nhỏ, cơ sở dữ liệu lớn và chiều cao.

Một số nội dung luận án đã nghiên cứu và giải quyết được như: (1) sử dụng thuộc tính thưa dòng để có thể loại đi được các đặc trưng dư thừa, giúp nâng cao độ chính xác tra cứu ảnh cho dù cỡ lớp của tập huấn luyện có thể là rất nhỏ; (2) đưa ra một mô hình mềm dẻo, có thể lựa chọn tập đặc trưng quan trọng, tự động bổ sung mẫu dương vào tập huấn luyện và không đòi hỏi số các mẫu dương phải đủ lớn; (3) tận dụng mô hình mạng nơ ron tích chập sâu autoencoder để học các biểu diễn đặc trưng hiệu quả cho tra cứu ảnh thông qua việc sử dụng shortcut connections trong kiến trúc autoencoder; (4) thiết kế một cơ chế học phản hồi liên quan sử dụng máy véc tơ hỗ trợ SVM để tận dụng các mẫu có nhãn từ phản hồi của người dùng.

Mặc dù luận án đã đạt được một số kết quả nghiên cứu quan trọng về lý luận khoa học và thực tiễn trong sử dụng kỹ thuật học máy vào quá trình CBIR với phản hồi liên quan, nhưng luận án vẫn còn một số vấn đề cần nghiên cứu, cải tiến và phát triển tiếp trong tương lai như: (1) Tận dụng các thành tựu của học máy hiện đại như mô hình Vision Transformer, mạng nơ ron tích chập đồ thị và cơ chế học truyền để nâng cao hiệu năng tra cứu; (2) Triển khai các giải pháp đã đề xuất vào việc giải quyết các lớp bài toán thực tiễn, có sử dụng dữ liệu hình ảnh với độ chính xác cao, thuộc nhiều lĩnh vực khác nhau như quân sự, y học, giáo dục, dự báo thời tiết, ….

NHỮNG ĐÓNG GÓP MỚI CỦA LUẬN ÁN

Luận án đã có hai đóng góp mới cho bài toán tra cứu ảnh như: phương pháp **SDAIR** *(Sparse Discriminant Analysis for Image Retrieval)* [CT4, CT2] và phương pháp **AIR** *(Autoencoders for Image Retrieval)* [CT1, CT3], cụ thể như sau:

- Đóng góp 1: Đề xuất được phương pháp tra cứu ảnh SDAIR. Phương pháp này kết hợp mô hình trích rút đặc trưng quan trọng dựa trên phương pháp RSLDA với mô hình phân lớp trong hệ thống CBIR nhằm cải tiến độ chính xác và thời gian truy vấn. SDAIR giải quyết được ba vấn đề: (1) Số các phản hồi (các mẫu) của người dùng quá nhỏ so với chiều của không gian đặc trưng; (2) Số các mẫu phản hồi dương ít hơn số các mẫu phản hồi âm rất nhiều; (3) Số lớp quá nhỏ, dẫn đến số các hướng chiếu bị giới hạn bởi số các lớp.

- Đóng góp 2: Đề xuất phương pháp tra cứu ảnh AIR dựa trên 3 thành phần: Huấn luyện bán giám sát bằng mạng nơ ron tích chập autoencoder, trích rút đặc trưng ảnh và phân lớp SVM trong phản hồi liên quan nhằm cải tiến độ chính xác và thời gian truy vấn. Mạng nơ ron tích chập autoencoder được tận dụng để học các biểu diễn đặc trưng hiệu quả cho tra cứu ảnh thông qua việc sử dụng shortcut connections trong kiến trúc autoencoder. Mô hình học này được sử dụng vào việc tạo ra các biểu diễn đặc trưng của ảnh cơ sở dữ liệu. Trên cơ sở các biểu diễn đặc trưng này, luận án đã thiết kế một cơ chế học RF sử dụng máy véc tơ hỗ trợ SVM để tận dụng các mẫu có nhãn từ phản hồi của người dùng. AIR giải quyết được hai hạn chế: (1) Khả năng phân biệt kém của các phương pháp đã có; (2) Giảm nhẹ vấn đề vanishing/exploding gradients và quá trình hội tụ nhanh.

DANH MỤC CÁC CÔNG TRÌNH CỦA TÁC GIẢ

1. An Hong Son, Nguyen Huu Quynh, Dao Thi Thuy Quynh, Cu Viet Dung, "Deep Learning of Image Representations with Convolutional Neural Networks Autoencoder for Image Retrieval with Relevance Feedback", *Journal on Information Technologies & Communications*, Vol. 2023, No. 1, pp. 17-24 (ISSN: 1859-3534, DOI: https://https://doi.org/10.32913/mic-ict-research.v2023.n1.1063).

2. Son An Hong, Quynh Dao Thi Thuy, Quynh Nguyen Huu, "Stuck Query Point Processing Of Multi-point Query For Image Retrieval With Relevance Feedback", *Journal of Information Hiding and Multimedia Signal Processing*, Vol. 12, No. 2, pp. 42-55, June 2021. (ISSN:2073-4212/2073-4239; **SCOPUS**).

3. Son An Hong, Quynh Nguyen Huu, Dung Cu Viet, Quynh Dao Thi Thuy, Tao Ngo Quoc (Accepted 20/5/2022), "Learning Binary Codes for Fast Image Retrieval with Sparse Discriminant Analysis and Deep Autoencoders", *Intelligent Data Analysis*, Vol. 27, No. 3, April 2023 (ISSN: 1088-467X/1571-4128; **SCIE**).

4. Son An Hong, Quynh Nguyen Huu, Dung Cu Viet, Quynh Dao Thi Thuy, Tao Ngo Quoc, "Improving image retrieval effectiveness via sparse discriminant analysis", *Multimedia Tools and Applications*, March 2023, pp.1-24 (ISSN: 1380-7501/1573-7721; **SCIE**).