BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

TRẦN VĂN HUY

KẾT HỢP XẾP HẠNG ĐA TẠP VÀ HỌC ĐỘ ĐO TƯƠNG TỰ CHO TRA CỨU ẢNH

LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

Hà Nội, năm 2025

BỘ CÔNG THƯƠNG
TRƯỜNG ĐẠI HỌC ĐIỆN LỰC

TRẦN VĂN HUY

KẾT HỢP XẾP HẠNG ĐA TẠP VÀ HỌC ĐỘ ĐO TƯƠNG TỰ CHO TRA CỨU ẢNH

Ngành: Công nghệ thông tin

Mã số: 9480201

LUẬN ÁN TIẾN SĨ CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC:
1. TS. NGÔ HOÀNG HUY
2. TS. NGUYỄN VĂN ĐOÀN

Hà Nội, năm 2025

LỜI CAM ĐOAN

Tôi xin cam đoan luận án tiến sĩ là kết quả nghiên cứu khoa học của tôi dưới sự hướng dẫn của TS. Ngô Hoàng Huy và TS. Nguyễn Văn Đoàn. Các kết quả nghiên cứu được trình bày trong luận án là trung thực, khách quan và chưa từng được tác giả khác công bố.

Tôi xin cam đoan rằng mọi sự giúp đỡ cho việc thực hiện luận án đã được cảm ơn, các thông tin trích dẫn trong luận án này đều được chỉ rõ nguồn gốc.

Hà Nội, ngày tháng 3 năm 2025

Tập thể hướng dẫn

Người hướng dẫn 1	**Người hướng dẫn 2**	**Tác giả luận án**
TS. Ngô Hoàng Huy	**TS. Nguyễn Văn Đoàn**	**Trần Văn Huy**

LỜI CẢM ƠN

Với tình cảm chân thành và lòng biết ơn sâu sắc, tôi xin trân trọng gửi lời cảm ơn tới Ban Lãnh đạo Đại học Điện lực cùng các thầy cô giáo tham gia giảng dạy đã cung cấp những kiến thức cơ bản, chuyên môn sâu và đã giúp đỡ tôi trong quá trình học tập nghiên cứu.

Đặc biệt tôi xin bày tỏ lòng biết ơn sâu sắc đến TS. Ngô Hoàng Huy và TS. Nguyễn Văn Đoàn những người hướng dẫn khoa học đã tận tâm giúp đỡ và chỉ dẫn cho tôi những kiến thức cũng như phương pháp luận trong suốt thời gian hướng dẫn nghiên cứu, hoàn thành luận án.

Xin cảm ơn Ban Giám hiệu Trường Đại học Hồng Đức, các đồng nghiệp tại Trung tâm Công nghệ thông tin và truyền thông - Trường Đại học Hồng Đức đã luôn động viên giúp đỡ tôi trong công tác để tôi có thời gian tập trung nghiên cứu và thực hiện luận án.

Đặc biệt tôi xin bày tỏ lòng biết ơn sâu sắc tới Cha, Mẹ, Vợ, Con và các anh, chị em trong gia đình, những người luôn dành cho tôi những tình cảm nồng ấm và sẻ chia những lúc khó khăn trong cuộc sống, luôn động viên giúp đỡ tôi trong quá trình nghiên cứu. Luận án cũng là món quà tinh thần mà tôi trân trọng gửi tặng đến các thành viên trong Gia đình.

Tôi xin trân trọng cảm ơn!

MỤC LỤC

DANH MỤC CÁC TỪ VIẾT TẮT VÀ KÝ HIỆU

Từ viết tắt	**Tên đầy đủ** (và tạm dịch)
CBIR	Tra cứu ảnh dựa trên nội dung
ARP	Độ chính xác trung bình (Averge Retrieval Precision)
MAP	Mean Average Precision
MinP	Minimum Precision
σP	Standard deviation of Precision
CSDL	Cơ sở dữ liệu
Precision	Độ chính xác
Recall	Độ triệu hồi
RF	Relevance feedback (Phản hồi liên quan)
FCM	Fuzzy C-mean (Thuật toán phân cụm mờ C-mean)
K-means	K-means (Thuật toán phân cụm K-means)
KNN	K-Nearest Neighbor (K- láng giềng gần nhất)
DML	Distance metric learning
CNN	Convolutional Neural Network
SVM	Support Vector Machine
SVC	Support Vector Classifier
SVR	Support Vector Regression
MR	Manifold Ranking
EMR	Efficient Manifold Ranking - Xếp hạng đa tạp hiệu quả
ERR	Eros rate Tỷ lệ lỗi trung bình
AGR	Anchor Graph Regulrization
EAGR	Efficient Anchor Graph Regulrization
SGR	Sub-Graph Regularization
NCA	Neighbourhood Components Analysis
DMLMJ	Distance Metric Learning through Maximization of the Jeffrey divergence
CMMML	Clustered Multi-Metric Learning
MAE	Độ sai số trung bình tuyệt đối (MAE - Mean Absolute Error)
RSME	Phương sai trung bình (RMSE - Root Mean Square Error)
R	Chỉ số tương quan (R - Correlation Coefficient),
CoEMR	Kết hợp xếp hạng đa tạp hiệu quả
LLMs	Large Language Models

Ký hiệu

Ký hiệu	Tên đầy đủ
I	Ảnh đa kênh nói chung
Q	Ảnh truy vấn
F	Không gian đặc trưng
E	CSDL ảnh
n	Số lượng ảnh của cơ sở dữ liệu ảnh *E*
FE	Đặc trưng ảnh
LF	Đặc trưng mức thấp của ảnh (Low-level Feature)
HF	Đặc trưng mức cao của ảnh (Hight-level Feature)
d	Số chiều của vector đặc trưng
DFE	Đặc trưng nhúng
CNNFE	Đặc trưng CNN
FE_i	Đặc trưng thô mức thấp của ảnh thứ *i*
LF_Q	Đặc trưng mức thấp của ảnh ảnh truy vấn.
HF_Q	Đặc trưng mức cao của ảnh ảnh truy vấn.
LF_Q_t	Vector đặc trưng mức thấp thuộc bộ *t* của ảnh truy vấn *Q*
$LF\_Q_t^{norm}$	Dữ liệu thuộc bộ *t,* đã chuẩn hoá của ảnh truy vấn
C	Số cụm trong thuật toán FCM (số Anchor point)
$d(E_i, A_l)$	Khoảng cách giữa 2 vector A và B có cùng số chiều.
r	Số lân cận trong thuật toán EMR
IC	Tập cặp véc tơ
VDR	Tập véc tơ sai khác - Vector different ranking
S	Độ đo tương tự xây dựng từ tập VDR
LLMs	Các mô hình ngôn ngữ lớn - Large language Models

DANH MỤC BẢNG

DANH MỤC HÌNH VẼ

MỞ ĐẦU

1. Tính cấp thiết của đề tài

Cơ sở dữ liệu ảnh ngày càng trở nên phổ biến trong các lĩnh vực ứng dụng khác nhau như y học, viễn thám, thời trang, phòng chống tội phạm, xuất bản, kiến trúc,… cùng với sự bùng nổ thông tin trên Internet, sự phát triển vượt trội của các công nghệ kỹ thuật số và sự phổ biến rộng rãi các thiết bị công nghệ quay phim, chụp ảnh dẫn đến kho dữ liệu ảnh lưu trữ và chia sẻ trên Internet cũng tăng lên nhanh chóng. Trong thập kỷ qua DOMO[1] thống kê dữ liệu của thế giới, cho thấy sự gia tăng đáng kể trong hoạt động trên internet, từ Instagram và X đến Amazon. Từ năm 2013 đến 2024 số người dùng Internet phát triển từ 2.1 tỷ người đến 5.52 tỷ người cùng với đó số lượng dữ liệu ảnh khổng lồ đã được tải lên internet. Một số cơ sở dữ liệu ảnh khổng lồ điển hình như Facebook, YouTube, Google,… rất lớn và đa dạng đã thúc đẩy sự quan tâm nghiên cứu các phương pháp khai thác hiệu quả cơ sở dữ liệu ảnh lớn này. Đây là một thách thức cho việc tổ chức và tìm kiếm ảnh theo cách truyền thống. Do đó, tra cứu hình ảnh tương tự từ các tập dữ liệu ảnh lớn là một bài toán quan trọng trong lĩnh vực thị giác máy tính [1]. Các kết quả khảo sát và dự báo của các nghiên cứu gần đây cho thấy việc tra cứu các hình ảnh liên quan với yêu cầu người dùng là bài toán phù hợp với nhu cầu xã hội hiện đại.

Để giải quyết bài toán tra cứu dữ liệu ảnh, đã có nhiều hệ thống tra cứu ảnh dựa trên nội dung (Content-Based Image Retrieval – CBIR) được phát triển như QBIC, Photobook, Visual-Seek, MARS, El Nino, CIRES, PicSOM, PicHunter, MIRROR, Virage, Netra, SIMPLITcity… [2], [3], tuy nhiên các hệ thống thống này cũng chưa đáp ứng được kỳ vọng của người dùng. Các công ty công nghệ lớn như Google, Facebook, Taobao, Bing, ByteDance… có riêng

[1] https://www.domo.com/learn/infographic/data-never-sleeps-12

các nhóm nghiên cứu các ứng dụng về tra cứu ảnh để phục vụ cho mục đích thương mại.

1.1. Tra cứu ảnh dựa trên nội dung

Kỹ thuật tra cứu ảnh dựa trên nội dung đã được phát triển để tìm kiếm các hình ảnh có liên quan từ cơ sở dữ liệu dựa trên đối tượng hoặc nội dung của hình ảnh đầu vào. Đây là một bài toán được áp dụng rộng rãi trong lĩnh vực thị giác máy tính và mang lại hiệu quả kinh tế trong nhiều ứng dụng, chẳng hạn như: tìm kiếm khuôn mặt, vân tay, hình ảnh y tế, kỹ thuật hình sự, thương mại điện tử và nhiều ứng dụng khác.

Các hệ thống CBIR có độ chính xác phụ thuộc vào hai yếu tố: (1) Các đặc trưng biểu diễn nội dung ảnh; (2) Các phương pháp tra cứu và xếp hạng theo ảnh truy vấn.

Các yếu tố này tương ứng với 2 pha của hệ thống CBIR: ở pha ngoại tuyến (*offline*) là trích rút các đặc trưng ảnh trong cơ sở dữ liệu để tìm ra một biểu diễn phù hợp dưới dạng vectơ đặc trưng *d* chiều; ở pha trực tuyến (*online*) là so khớp vector đặc trưng ảnh truy vấn với cơ sở dữ liệu vector đặc trưng ảnh, sắp xếp thứ hạng các ảnh có độ tương tự cao nhất (hoặc là sắp xếp thứ tự ảnh theo thứ tự từ thấp đến cao của độ đo "không tương tự" như các độ đo khoảng cách) và sau đó trả về tập ảnh kết quả.

Các đặc trưng được sử dụng để biểu diễn nội dung ảnh: Trích rút đặc trưng và biểu diễn ảnh trong pha ngoại tuyến là giai đoạn cơ bản và quan trọng trong các hệ thống CBIR. Zheng và các cộng sự [1] đã chỉ ra nếu ảnh được biểu diễn bởi các đặc trưng "tốt" có thể giúp cải thiện độ chính xác của tìm kiếm tương tự lên tới 51,3%. Trong nhiều thập kỷ qua, một loạt các kỹ thuật trích rút đặc trưng được đề xuất và nghiên cứu để tìm ra các biểu diễn hình ảnh phong phú và đầy đủ hơn về mặt ngữ nghĩa, nhưng nó vẫn còn là một thách thức to lớn trong các ứng dụng CBIR. Để thu hẹp khoảng cách ngữ nghĩa [3], việc sử

dụng mạng học sâu để trích xuất đặc trưng ảnh đã mang lại hiệu quả đáng kể. Trong những năm gần đây, với sự phát triển mạnh mẽ của mạng học sâu (Deep Learning - DL) [3], các hệ thống CBIR dựa trên học sâu đã được triển khai và đạt được hiệu quả cao [4, 5]. Các mạng CNN như VGG, ResNet, Inception và EfficientNet để trích xuất đặc trưng biểu diễn ảnh đã đem lại hiệu suất đáng kể cho các hệ thống CBIR [15, 16, 17]. Đối với các đặc trưng CNN, trong công trình [12] các tác giả đã trình bày một phân tích và thử nghiệm về bộ dữ liệu ImageNet, từ đó các tác giả chỉ ra rằng hai biểu diễn đặc trưng fc4096a và fc4096b, được trích xuất từ lớp thứ nhất và lớp thứ hai của AlexNet có tính khái quát tốt hơn khả năng hơn các đặc trưng khác của CNN và mang lại hiệu suất cao cho tra cứu ảnh.

Các phương pháp tra cứu và xếp hạng ảnh: Các hệ thống tra cứu ảnh sử dụng các độ đo tương tự hiệu quả để so khớp các đặc trưng nội dung của ảnh truy vấn với đặc trưng có trong cơ sở dữ liệu hình ảnh. Ngay với cả những đặc trưng "tốt" thì các độ đo tương tự vẫn đóng vai trò quan trọng để tăng được độ chính xác. Trong các nghiên cứu [6, 7, 8] cho thấy, các hệ thống CBIR sử dụng các đặc trưng CNN đều dùng độ đo khoảng cách “truyền thống” (ví dụ: Khoảng cách Euclide) để đánh giá sự giống nhau của hai hình ảnh dẫn đến hiệu suất của tìm kiếm độ tương tự không như mong muốn [9]. Các độ đo truyền thống chỉ nắm bắt được sự tương đồng giữa hai hình ảnh ở mức cục bộ mà không xem xét độ tương tự giữa một nhóm các hình ảnh tương tự ở mức toàn cục. Ngoài ra, cần nhấn mạnh rằng các kỹ thuật xếp hạng dựa trên cấu trúc nội tại của cơ sở dữ liệu ảnh cũng rất quan trọng trong quá trình tra cứu và xếp hạng ảnh. Việc sử dụng kỹ thuật xếp hạng dựa trên cấu trúc nội tại của CSDL giúp phân loại và xếp hạng các ảnh dựa trên mức độ tương tự và không tương tự với ảnh truy vấn. Bằng cách ưu tiên các kết quả tìm kiếm sao cho ảnh có độ tương tự cao với ảnh truy vấn được hiển thị đầu tiên, kỹ thuật xếp hạng dựa trên cấu trúc nội tại của CSDL cải thiện tính toàn vẹn và khả năng phân loại của hệ thống tra

cứu ảnh. Điều này giúp đảm bảo kết quả trả về chính xác và phù hợp với yêu cầu của người dùng trong quá trình tìm kiếm ảnh.

1.2. Tra cứu ảnh, xếp hạng tương tự ảnh và xếp hạng đa tạp.

Trong các hệ thống CBIR, để ước lượng độ tương đồng giữa hai vector đặc trưng (biểu diễn hai hình ảnh cần đối sánh), các kỹ thuật ước lượng về độ tương tự, về khoảng cách và xếp hạng đa tạp được sử dụng.

Các hệ thống tra cứu ảnh thường sử dụng các độ đo tương tự hoặc độ đo không tương tự (như độ đo khoảng cách) để so sánh các đặc trưng nội dung của ảnh truy vấn với các đặc trưng có trong cơ sở dữ liệu hình ảnh. Ngay cả khi sử dụng các đặc trưng có độ phân biệt ảnh cao, các độ đo tương tự vẫn đóng vai trò quan trọng để nâng cao độ chính xác. Tuy nhiên, trong nghiên cứu đã chỉ ra rằng các hệ thống CBIR sử dụng các đặc trưng CNN thường sử dụng các độ đo khoảng cách "truyền thống", các độ đo truyền thống thường chỉ phản ánh sự tương đồng giữa hai hình ảnh ở mức cục bộ mà không xem xét sự tương tự giữa một nhóm các hình ảnh tương tự ở mức toàn cục như khoảng cách Euclid để đánh giá sự tương tự giữa hai hình ảnh, và điều này thường dẫn đến hiệu suất tra cứu ảnh không như mong đợi [18 - 21]

Để khắc phục những hạn chế này và để khám phá cấu trúc phi tuyến của dữ liệu đặc trưng ảnh, các phương pháp xếp hạng đa tạp trong CBIR đã được đề xuất [22, 23]. Phương pháp xếp hạng đa tạp (Manifold Ranking-MR) nhằm khám phá cấu trúc phi tuyến của dữ liệu bằng cách xem xét các mẫu dữ liệu trên nhiều không gian con khác nhau [21,22,24]. Trong các nghiên cứu [36, 47-50] đã chứng minh được hiệu quả của MR trong CBIR với biểu diễn ảnh bằng kết hợp các đặc trưng mức thấp và MR được coi là một phương pháp hiệu quả để nắm bắt sự tương đồng không những ở cục bộ giữa các cặp điểm dữ liệu mà còn dựa trên cấu trúc tổng thể của toàn bộ dữ liệu. Thay vì chỉ xem xét độ tương tự giữa từng cặp điểm dữ liệu, MR xem xét toàn bộ không gian dữ liệu để đưa

ra điểm số xếp hạng. Điều này giúp xếp hạng các mẫu dữ liệu theo một cách có ý nghĩa hơn, dựa trên mức độ liên quan ngữ nghĩa tổng thể, điều này rất hữu ích trong các ứng dụng như tìm kiếm dựa trên nội dung, nơi mà mục đích là tìm ra các mẫu dữ liệu có liên quan ngữ nghĩa cao.

Do khả năng đa dạng của MR, trong những năm gần đây phương pháp này được ứng dụng trong cho nhiều lĩnh vực như tái nhận dạng người [91], tìm kiếm sự tương đồng của tài liệu [92], xác định mối quan hệ hóa học định lượng [93], phát phát hiện điểm nổi trội trong ảnh và tra cứu ảnh [79, 94, 95]. Có nhiều phương pháp xếp hạng đa tạp trong CBIR, như Tra cứu ảnh dựa vào nội dung với xếp hạng đa tạp nhanh (Fast Manifold-Ranking for Content-Based Image Retrieval - FMR) [25], Tra cứu ảnh dựa vào xếp hạng đa tạp mở rộng (Scaling Manifold Ranking Based Image Retrieval - SMR) [26], tra cứu ảnh với xếp hạng đa tạp hiệu quả (Efficient Manifold Ranking for Image Retrieval - EMR) [42], tra cứu ảnh dựa vào nội dung với mô hình xếp hạng dựa trên đồ thị mở rộng (EMR: A Scalable Graph-Based Ranking Model for Content-Based Image Retrieval) [36], Tăng cường phản hồi liên quan dài hạn trong CBIR với tối ưu hóa mở rộng đồ thị con (SGR - A scalable sub-graph regularization for efficient content based image retrieval with long-term relevance feedback enhancement) [39]… Trong đó, tra cứu ảnh với xếp hạng đa tạp hiệu quả là phương pháp xếp hạng đa tạp hiệu quả được Bin Xu và các cộng sự đề xuất [42]. Phương pháp này tập trung giải quyết hai vấn đề chính:

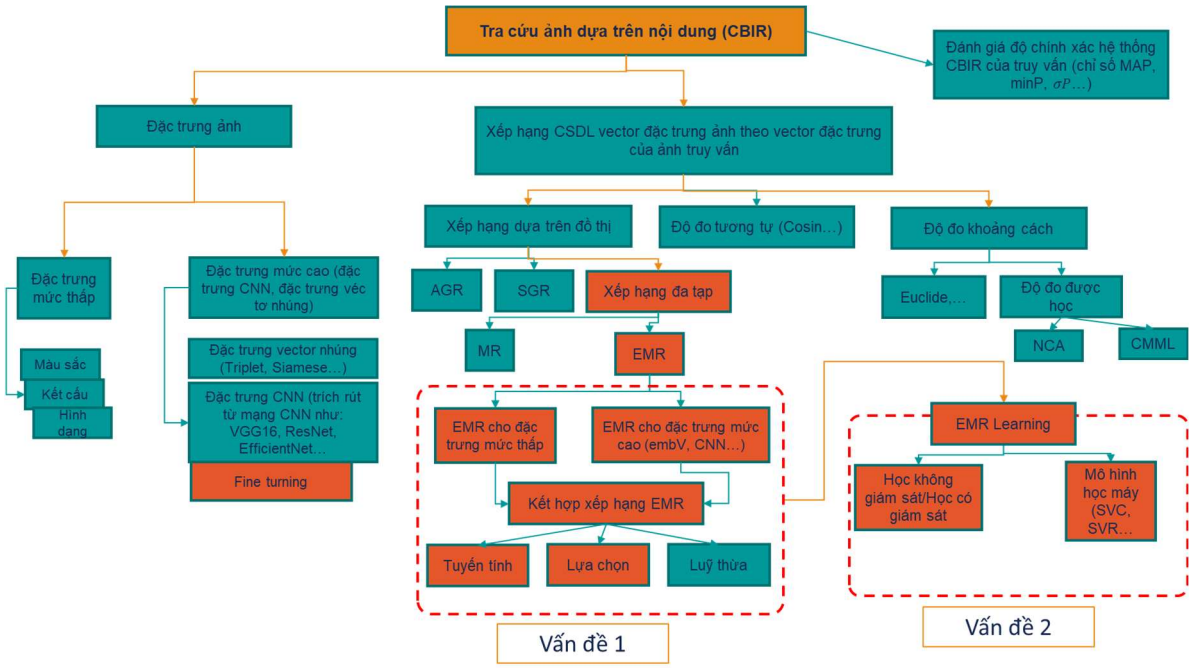
1- Xây dựng đồ thị neo có khả năng mở rộng thay cho đồ thị *K-NN* truyền thống.

2- Xây dựng mô hình tính toán xếp hạng hiệu quả, có chi phí tính toán nhỏ bằng cách thiết kế hình thức mới của ma trận kề để có thể xử lý tra cứu theo thời gian thực trên CSDL lớn.

Trong [122, 128] các tác giả đã sử dụng cách kết hợp nhiều bộ đặc trưng khác nhau như đặc trưng mức thấp (Color, Shape, GIST, Texture, LBP) và đặc

trưng CNN (EfficientNET), sau đó sử dụng xếp hạng đa tạp trên các bộ đặc trưng kết hợp (tuyến tính) thành một xếp hạng EMR cho bộ véc tơ đặc trưng với số chiều rất cao (>2048). Các phương pháp này chủ yếu tập trung vào việc sử dụng nhiều đặc trưng ảnh trong xếp hạng EMR để cải thiện hiệu quả tra cứu ảnh. Mặc dù cách tiếp cận này khá hiệu quả, nhưng do chỉ sử dụng một bộ xếp hạng duy nhất, chúng vẫn tồn tại một số hạn chế. Cụ thể, phương pháp này chưa khai thác tối đa ưu điểm của từng loại đặc trưng ảnh, thiếu sự linh hoạt trong việc điều chỉnh tham số để nâng cao độ chính xác xếp hạng, đồng thời làm tăng độ phức tạp tính toán do véc tơ đặc trưng có số chiều rất lớn.

Để làm rõ hơn nội dung vấn đề nghiên cứu, luận án mô hình hóa và định vị vấn đề, vùng nghiên cứu trong CBIR cho các đề xuất của luận án trong **Hình 0.1** được mô tả như sau:



Hình 0.1: Định vị vùng nghiên cứu trong CBIR của luận án

1.3. Kết luận khảo sát và đề xuất hướng nghiên cứu

EMR mặc dù là một phương pháp mạnh mẽ, nhưng vẫn tồn tại các vấn đề như sau:

1) Tính toán chậm với số lượng điểm neo (anchor) lớn.

EMR cần sử dụng các điểm tham chiếu, được gọi là anchor, để tính toán độ tương tự giữa các hình ảnh. Với các cơ sở dữ liệu hình ảnh lớn, việc sử dụng nhiều anchor có thể dẫn đến hiệu suất tính toán kém. Các phép toán ma trận trên các ma trận lớn có thể đòi hỏi nhiều thời gian và tài nguyên tính toán.

2) Có hạn chế khi thêm ảnh mới vào cơ sở dữ liệu.

EMR yêu cầu xây dựng lại mô hình hoặc tính toán lại các ma trận xác định đồ thị kề dù có thể vẫn giữ nguyên tập điểm neo khi thêm ảnh mới vào cơ sở dữ liệu. Điều này có thể là một quá trình tốn thời gian và tài nguyên, đặc biệt là trong các hệ thống quản lý cơ sở dữ liệu động hoặc cần mở rộng.

3) Hạn chế trong việc xác định độ tương tự với ảnh ngoài cơ sở dữ liệu.

EMR chỉ cho phép tính toán độ tương tự giữa ảnh truy vấn và các ảnh trong cơ sở dữ liệu. Điều này đồng nghĩa với việc không thể xác định độ tương tự giữa hai ảnh nằm ngoài cơ sở dữ liệu. Vì vậy việc ứng dụng của phương pháp này trong các tác vụ liên quan đến phát hiện và phân loại hình ảnh trong thực tế còn hạn chế.

Ban đầu, EMR được phát triển để thực hiện nhiệm vụ trong môi trường không giám sát, có nghĩa là không cần có dữ liệu nhãn về mức độ tương tự giữa các cặp hình ảnh trong cơ sở dữ liệu. EMR sử dụng thông tin cơ bản về mức độ tương đồng giữa các hình ảnh để xếp hạng chúng theo một thứ tự. Tuy nhiên, để có thể xử lý hiệu quả các tình huống phức tạp hơn, như xác định độ tương tự giữa các hình ảnh không có trong cơ sở dữ liệu, chúng ta cần phải tận dụng sức mạnh của học máy. Học bán giám sát dựa trên giá trị xếp hạng của EMR sử dụng các mô hình học máy, thường là các mạng nơ-ron sâu hoặc các thuật toán học máy tiên tiến, để tự động xác định độ tương tự giữa các cặp hình ảnh. Mô hình này sẽ sử dụng thông tin từ EMR, như các giá trị xếp hạng đã được

tạo ra, để làm dữ liệu đầu vào cho quá trình học và cải thiện khả năng xác định độ tương tự. Khi kết hợp EMR và học bán giám sát dựa trên giá trị xếp hạng của EMR, chúng ta có khả năng mở rộng ứng dụng của EMR. Chúng ta không chỉ sử dụng EMR để tìm kiếm hình ảnh trong cơ sở dữ liệu, mà còn để đánh giá độ tương tự với các hình ảnh không có trong cơ sở dữ liệu. Điều này mang lại sự linh hoạt và khả năng tùy chỉnh để giải quyết các tình huống phức tạp hơn trong lĩnh vực xử lý hình ảnh và truy vấn dữ liệu.

Dựa trên những phân tích tổng quan về CBIR, có thể nhận định rằng đây là một công nghệ quan trọng và cần thiết trong bối cảnh dữ liệu hình ảnh ngày càng gia tăng. Những nghiên cứu mới nhất, đặc biệt là việc ứng dụng xếp hạng đa tạp hiệu quả trong CBIR, đã mở ra nhiều hướng đi mới, giúp cải thiện độ chính xác và hiệu suất tra cứu ảnh. Tuy nhiên, vẫn còn nhiều thách thức khi sử dụng EMR, như: độ phức tạp trong việc xây dựng mô hình, yêu cầu tài nguyên tính toán cao và khó khăn trong việc so sánh độ tương tự với những hình ảnh nằm ngoài CSDL. Để giải quyết một phần các hạn chế này, luận án đã chọn đề tài - ***Kết hợp xếp hạng đa tạp và học độ đo tương tự cho tra cứu ảnh.***

2. Mục tiêu của luận án

Mục tiêu chung của luận án: Nâng cao hiệu quả tra cứu ảnh dựa trên kết hợp xếp hạng đa tạp và tiếp cận học độ đo tương tự.

Mục tiêu cụ thể của luận án:

Nghiên cứu, đề xuất một số giải pháp nâng cao độ chính xác tra cứu ảnh dựa trên nội dung theo tiếp cận xếp hạng đa tạp bao gồm:

- Nghiên cứu thuật toán xếp hạng đa tạp hiệu quả, đề xuất kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp với xếp hạng của hình ảnh đặc trưng mức cao.

- Nghiên cứu, đề xuất xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR để giải quyết việc ảnh tra cứu nằm ngoài cơ

sở dữ liệu.

3. Đối tượng nghiên cứu của luận án

Luận án tập trung vào nghiên cứu và tìm hiểu một số đối tượng liên quan đến tra cứu ảnh như:

- Tổng quan về Tra cứu ảnh dựa vào nội dung.

- Kỹ thuật biểu diễn ảnh với đặc trưng mức thấp, đặc trưng mức cao gồm đặc trưng véc tơ nhúng, đặc trưng CNN.

- Xếp hạng đa tạp trong tra cứu ảnh dựa vào nội dung.

- Các kỹ thuật học máy, học bán giám sát các giá trị xếp hạng EMR.

- Môi trường thực nghiệm, tập dữ liệu ảnh thực nghiệm và phương pháp đánh giá độ chính xác.

4. Phạm vi nghiên cứu

Trong luận án này, phạm vi nghiên cứu bao gồm:

- Nghiên cứu thuật toán xếp hạng đa tạp trong tra cứu ảnh dựa vào nội dung.

- Nghiên cứu phương pháp tổ hợp xếp hạng đa tạp hiệu quả kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp với xếp hạng của hình ảnh đặc trưng mức cao (đặc trưng véc tơ nhúng, đặc trưng CNN).

- Nghiên cứu phương pháp nâng cao hiệu quả tra cứu ảnh bằng cách xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR. Đề xuất kỹ thuật đo độ tương tự ảnh EMR Learning để giải quyết việc ảnh tra cứu nằm ngoài cơ sở dữ liệu.

- Trong phạm vi của luận án chỉ tập trung nâng cao chất lượng tra cứu về độ chính xác, các vấn đề về thời gian cho một truy vấn cũng được xem xét ở khía cạnh có thể chấp nhận được.

5. Các đóng góp của luận án

Nhằm mục tiêu nâng cao độ chính xác của tra cứu ảnh sử dụng phương học độ đo tương tự, luận án có các đóng góp sau:

(1) Nghiên cứu phương pháp tổ hợp các bộ xếp hạng đa tạp hiệu quả, đề xuất thuật toán CoEMR kết hợp nhiều bộ xếp hạng hình ảnh theo đặc trưng mức thấp, xếp hạng của hình ảnh đặc trưng mức cao [CT1, CT2, CT3, CT4]. Đề xuất phương pháp sử dụng truy vấn nhiều ảnh trên CBIR [CT8].

Giả sử rằng tập dữ liệu gốc có chứa các hình ảnh, mỗi hình ảnh có một véc tơ nhúng duy nhất tương ứng với đối tượng lớn nhất được xác định bởi Deep Metric Learning (DML) [128], được biểu thị tương ứng $\{I_1, I_2, I_3 \dots\}$. Tùy thuộc vào giá trị của I_q, luận án chọn xếp hạng hình ảnh theo đặc trưng mức thấp $r^*_{lf.v.Q}$ hoặc kết hợp với xếp hạng của hình ảnh đặc trưng mức cao (trích rút từ mạng CNN), trong đó $r^*_{lf.Q.i}, r^*_{emb.Q.i}$ được tính bằng EMR. Sau đó tiến hành đánh giá trên một số bộ dữ liệu phổ biến như Corel, VGG60k, Leaf30, Sign300HD... và so sánh với một số phương pháp xếp hạng dựa trên đồ thị khác như AGR, SGR... để chứng minh hiệu quả của thuật toán đề xuất.

(2) Nghiên cứu phương pháp nâng cao hiệu quả tra cứu ảnh bằng cách xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR. Đề xuất thuật toán đo độ tương tự ảnh EMR Learning để giải quyết việc ảnh tra cứu nằm ngoài cơ sở dữ liệu [CT5, CT6, CT7] gồm các nội dung:

- Chuẩn hoá đặc trưng của ảnh truy vấn và các ảnh trong CSDL sau đó tiến hành xếp hạng các EMR. Song song với đó tạo ra tập véc tơ sai khác làm đầu vào cho mô hình học hồi quy.
- Xây dựng mô hình EMR Learning bằng hồi quy SVM hoặc Random Forest…
- Với cặp ảnh đầu vào I_1, I_2 trích chọn đặc trưng ảnh, chuẩn hóa các véc tơ tương ứng V_1, V_2 đưa vào mô hình EMR Learning thu được xếp hạng r_{21} -

Chính là giá trị độ tương tự của cặp ảnh I_1, I_2.

Sau đó tiến hành đánh giá trên một số bộ dữ liệu phổ biến như Corel, VGG60k, Leaf, Sign300HD... và so sánh với một số phương pháp học độ đo khoảng cách khác như NCA, CMML để chứng minh hiệu quả của thuật toán đề xuất.

6. Bố cục của luận án

Luận án được tổ chức thành ba chương:

Chương 1: Tra cứu ảnh dựa trên nội dung (CBIR).

Trong chương này giới thiệu tổng quan về tra cứu ảnh dựa vào nội dung, trình bày học độ đo khoảng cách, độ đo tương tự, tập dữ liệu ảnh thực nghiệm và cách thức đánh giá độ chính xác của hệ thống tra cứu ảnh.

Chương 2: Phương pháp tra cứu ảnh sử dụng thuật toán kết hợp nhiều bộ xếp hạng đa tạp hiệu quả.

Trong chương này, luận án đề xuất phương pháp kết hợp nhiều bộ xếp hạng EMR (CoEMR), đánh giá hiệu quả của thuật toán CoEMR đề xuất trên một số tập dữ liệu có dạng "đa tạp" với các chỉ số khác nhau. Trong Chương 2, luận án cũng tiến hành thực nghiệm và đưa ra kết quả khi tra cứu ảnh bằng hệ thống CBIR sử dụng CoEMR trên các tập dữ liệu VGG60K, Leaf30, Corel, SIGN300HD…

Ngoài ra, trong Chương 2 luận án cũng đề xuất phương pháp truy vấn nhiều ảnh trên CBIR làm rõ tính hiệu quả của phương pháp kết hợp các xếp hạng EMR trong CBIR.

Chương 3: Xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR.

Trong chương 3, luận án đề xuất phương pháp xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR (EMR Learning). Xây dựng qua 3 bước đánh giá:

Bước 1: Xây dựng độ đo tương tự EMR

Bước 2: Học độ đo tương tự EMR Learning

Bước 3: Tính toán độ tương tự ảnh

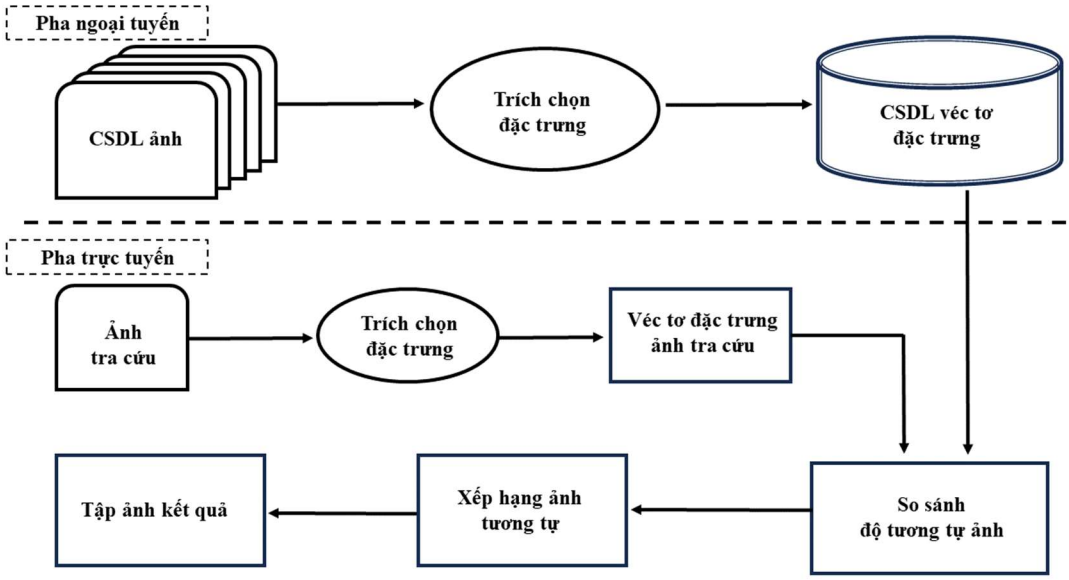
Thực hiện đánh giá hiệu quả của thuật toán EMR Learning trên các tập dữ liệu và so sánh với một số phương pháp học độ đo khoảng cách khác như NCA, CMML… Ngoài ra, luận án cũng đề xuất phương pháp sử dụng EMR cải tiến cho nhận dạng nhãn [CT5-CT7] nhằm giải quyết vấn đề đánh giá mức độ tương tự giữa hai ảnh đầu vào.

Cuối cùng, luận án đưa ra một số đề xuất và định hướng nghiên cứu trong tương lai.

CHƯƠNG 1: TRA CỨU ẢNH DỰA TRÊN NỘI DUNG

1.1 Giới thiệu về tra cứu ảnh dựa trên nội dung

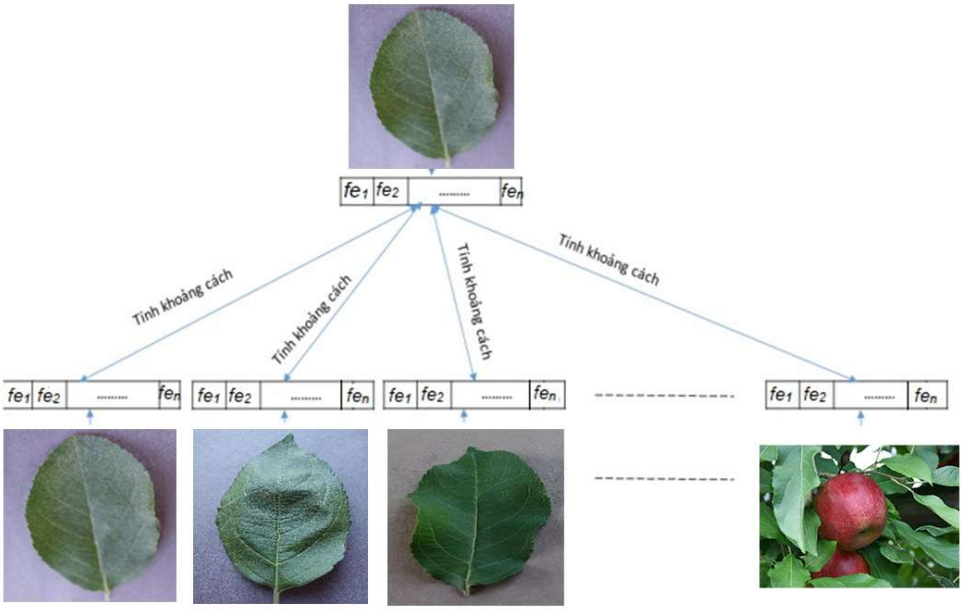
Tra cứu ảnh dựa vào nội dung (CBIR - Content based image retrieval) [2,10] thu hút rất nhiều sự chú ý từ các nhà nghiên cứu và được sử dụng nhiều trong công nghiệp, thương mại trong những năm qua do nhiều ứng dụng hữu ích của nó. Các thuật toán tra cứu ảnh thường xây dựng các độ đo tương tự toàn cục giữa các vector đặc trưng biểu diễn đối tượng ảnh đối sánh với toàn bộ vector đặc trưng trong CSDL.



Hình 1.1: Hệ thống tra cứu ảnh CBIR truyền thống

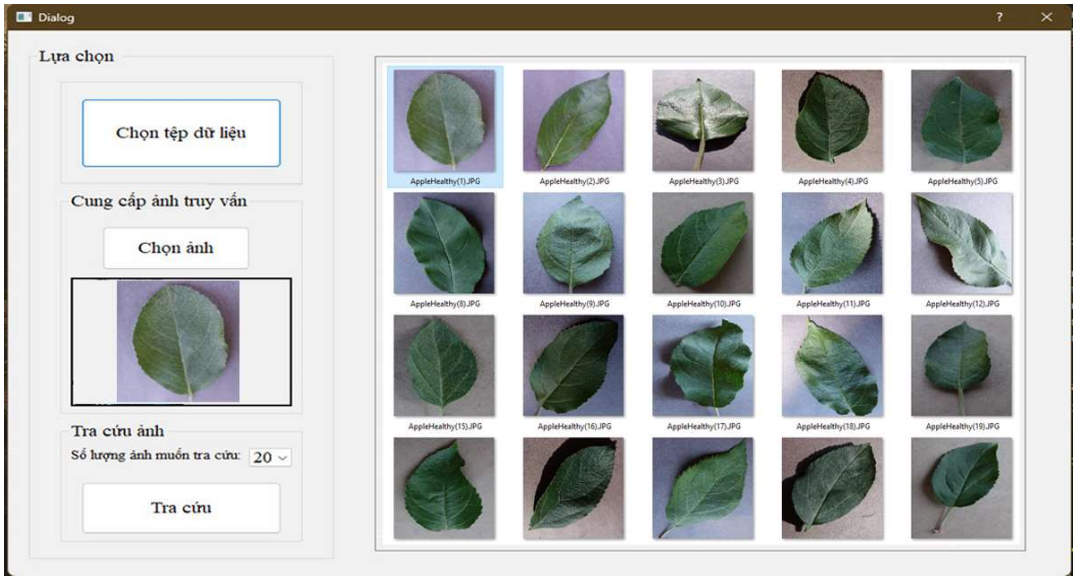
Hệ thống tra cứu ảnh CBIR như Hình 1.1 là kỹ thuật tra cứu ảnh được sử dụng để tìm ra tập các ảnh tương tự nhất đối với ảnh truy vấn mà người dùng đưa vào. Một hệ thống CBIR tiêu biểu được chia thành hai pha: trích rút đặc trưng ngoại tuyến và pha tra cứu ảnh trực tuyến. Trong pha ngoại tuyến, hệ thống trích rút tự động các thuộc tính trực quan ở mức thấp (LF) hoặc đặc trưng mức cao (HF) hoặc các loại đặc trưng được kết hợp với nhau. Trong pha tra cứu ảnh trực tuyến, người dùng cung cấp một ảnh mẫu cho hệ thống tra cứu để tìm kiếm các ảnh mong muốn (ở đây, độ dài véc tơ đặc trưng của ảnh tra cứu có cùng độ

dài với véc tơ đặc trưng của ảnh CSDL. Để trả lời tra cứu, hệ thống CBIR tìm trong CSDL ảnh để đưa ra những ảnh tương tự với ảnh truy vấn (như mô tả trong Hình 1.2). Cuối cùng hệ thống xếp hạng các ảnh theo thứ tự tăng dần của khoảng cách hay giảm dần của độ tương tự và trả về tập ảnh kết quả cho người dùng.



Hình 1.2. Minh họa đối sánh trong CBIR

Giao diện trực quan cho hệ thống CBIR truyền thống như Hình 1.3. Trong hình này người dùng cung cấp cho hệ thống ảnh tra cứu là hình ảnh của một lá cây táo khỏe mạnh, sau đó chúng ta thu được kết quả trả về bao gồm 20 ảnh.



Hình 1.3. Giao diện hệ thống CBIR truyền thống

Thông thường trong các hệ thống CBIR truyền thống một số đặc trưng được trích rút và thường sử dụng như đặc trưng màu sắc, kết cấu, hình dạng (vùng và đường viền).... Các đặc trưng này thường được chia thành hai nhóm: thứ nhất, nhóm đặc trưng toàn cục mô tả toàn bộ hình ảnh, nhóm còn lại là đặc trưng cục bộ, mà chia ảnh thành các vùng nhỏ hơn.

1.2 Đặc trưng ảnh

Đặc trưng ảnh là các thuộc tính được trích rút ra từ ảnh dựa trên phân phối cấu trúc, đối tượng có trong ảnh, từ đó tổng hợp các dữ liệu riêng lẻ để xác định thông tin có thể đo lường được khi quan sát, phân tích tình huống từ dữ liệu hình ảnh. Phương pháp trích rút đặc trưng và biểu diễn đặc trưng để thể hiện một cách riêng biệt và độc lập là yêu cầu quan trọng và quyết định mức độ thành công của các phương pháp nhận dạng mẫu, học máy, phân loại hay trong tra cứu ảnh…

Trong [11] trích rút đặc trưng ảnh được hiểu là quá trình biến đổi dữ liệu ảnh thô ban đầu quan sát được thành các tín hiệu hoặc dữ liệu có nhiều thông tin hơn để biểu diễn và xử lý hiệu quả hơn so với dữ liệu ảnh thô ban đầu, mang được nhiều ý nghĩa, giữ lại những thông tin quan trọng phục vụ cho việc phân tích và xử lý ngữ nghĩa hơn. Trích rút đặc trưng cho phép ánh xạ ảnh từ không gian ảnh sang không gian đặc trưng. Hiệu quả tra cứu ảnh phụ thuộc vào khả năng mô tả nội dung ảnh cho các ứng dụng cụ thể. Không tồn tại cách biểu diễn tốt nhất cho các đặc trưng thị giác vì mỗi đặc trưng có thể có nhiều cách biểu diễn theo các ngữ cảnh khác nhau. Chẳng hạn, đặc trưng màu có thể được biểu diễn bởi biểu đồ màu và mô men màu; đặc trưng hình dạng có thể biểu diễn bởi biểu đồ hệ số góc và GIST; đặc trưng kết cấu có thể biểu diễn bởi mã nhị phân cục bộ và biến đổi wavelet.

Trong thực tế, do ảnh đầu vào có thể được thu nhận và số hóa trong các điều kiện khác nhau (ánh sáng, môi trường, góc thu nhận,…) nên các đặc trưng

trích rút thường thỏa mãn một số tính chất bất biến như: Bất biến với phép tỉ lệ (scale invariance); bất biến với phép xoay (rotation invariance); bất biến với ánh sáng (intensity invariance); giàu thông tin; tính chính xác (accuracy); hiệu quả (efficiency); bền vững với nhiễu,… Một thuật toán (hay kỹ thuật) dùng để trích rút các đặc trưng ảnh được gọi là bộ trích rút đặc trưng (feature extraction).

Có nhiều phương pháp trích rút các đặc trưng trong CBIR không chỉ dựa trên toàn bộ ảnh mà thông qua các vùng được tách ra từ ảnh. Sharif và cộng sự [12] đề xuất một hệ thống CBIR phụ thuộc vào việc hợp nhất các từ trực quan (visual words) mà được tạo ra từ đặc trưng SIFT (scale invariant feature transform) và BRISK (binary robust invariant scalable keypoints). Yousuf và cộng sự [13] thực hiện một hệ thống CBIR dựa trên SFIT và LIOP (local intensity order pattern). LIOP đã được sử dụng để khắc phục hạn chế của SIFT trong việc thay đổi ánh sáng và các vùng có độ tương phản thấp. Việc sử dụng đặc trưng SIFT trong CBIR cho hiệu quả kém khi số chiều đặc trưng SIFT là rất lớn. Herbert và cộng sự [14] đề xuất đặc trưng SURF (speededup robust features) là một bộ mô tả cục bộ mạnh khác mà vượt qua giới hạn về số chiều cao của SIFT. SURF nhanh và mạnh hơn SIFT vì nó yêu cầu ít thời gian để tính toán và đối sánh các ảnh thông qua sử dụng cơ chế đánh chỉ số dựa trên tín hiệu Laplacian. Jabeen và cộng sự [15] đề xuất một hệ thống CBIR mới dựa trên việc kết hợp hai bộ mô tả SURF, FREAK (fast retina key point) để tạo thành các từ trực quan trên cơ sở của BoVW. Sau đó, phân cụm K-means được áp dụng trên các từ trực quan đó để tính toán một lược đồ cho các từ của mỗi ảnh.

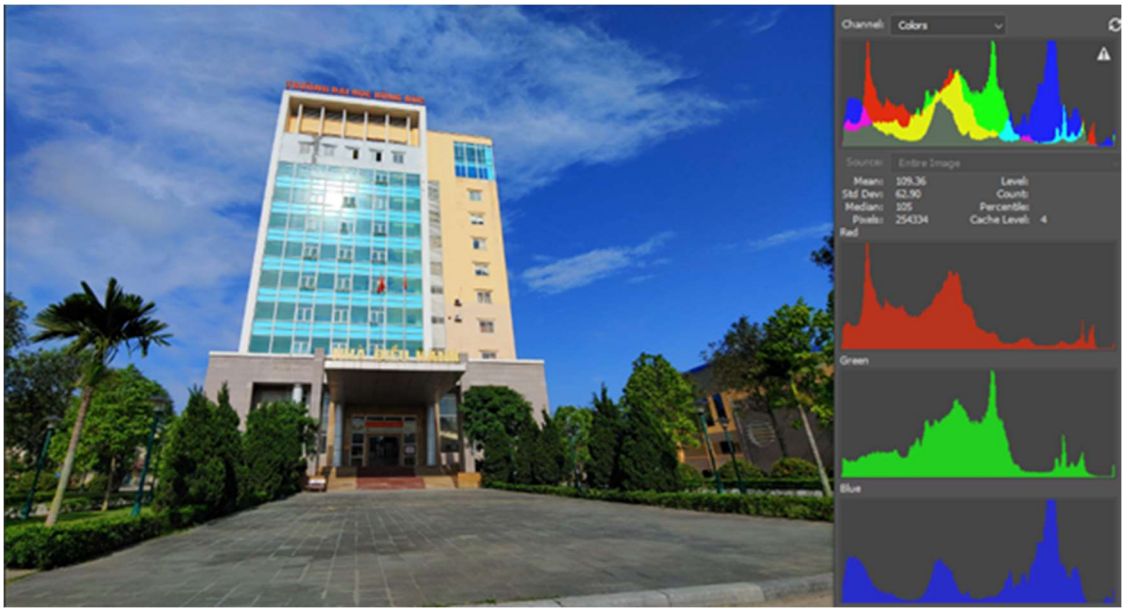
1.2.1 Đặc trưng mức thấp của ảnh

Đặc trưng màu sắc

Đặc trưng màu sắc được rất nhiều hệ thống tìm kiếm hình ảnh dựa trên nội dung nghiên cứu và sử dụng. Đặc trưng màu sắc ổn định và hầu như không bị ảnh hưởng bởi việc dịch chuyển, tỷ lệ và xoay hình ảnh. Đồng thời, màu sắc

có quan hệ với các đối tượng ảnh, nền, cho biết sự thay đổi trong vị trí, thời gian… Các biểu diễn mầu phổ biến là lược đồ mầu, mô men mầu, tương quan mầu và ma trận đồng hiện mầu.

Tùy mục đích của phương pháp truy vấn, các nhóm nghiên cứu có thể sử dụng các không gian màu khác nhau như phương pháp sử dụng không gian màu YCbCr với biểu đồ cạnh Canny và biến đổi Wavelet rời rạc [16], phương pháp sử dụng biểu đồ chênh lệch màu CDH (Color Difference Histogram) trong không gian màu HSV [17],… Biểu đồ màu (Color Histogram) [18] là sự mô tả sự biến đổi màu sắc trong một ảnh. Biểu đồ màu của ảnh bất biến với hướng và chỉ thay đổi dần theo góc nhìn. Tuy nhiên, biểu đồ màu không nắm bắt được mối quan hệ không gian của các vùng màu và khả năng phân biệt bị giới hạn. Do đó, mô men màu được sử dụng độ lệch chuẩn và giá trị trung bình của các phân phối trong mỗi dải màu cho mục đích lập chỉ mục màu trong các ứng dụng tra cứu ảnh để so sánh sự giống nhau về màu sắc giữa hai ảnh giống nhau [19, 20].



Hình 1.4: Mô tả biểu đồ màu của ảnh

Đặc trưng kết cấu

Kết cấu là một đặc trưng ảnh quan trọng để mô tả các thuộc tính bề mặt

của một đối tượng như độ mịn, độ thô, độ sâu, ... và mối quan hệ của nó với các vùng xung quanh như sự thay đổi độ sáng cục bộ trong một vùng lân cận, sự sắp xếp không gian của các mức xám,... [21]. Nhiều phương pháp tìm kiếm ảnh theo nội dung (CBIR) dựa trên kết cấu được đề xuất. Kỹ thuật phân tích kết cấu thống kê chủ yếu mô tả kết cấu của các vùng trong ảnh dựa vào biểu đồ mức xám [21]. Ma trận đồng xuất hiện mức xám GLCM (Gray-level co-occurrence matrix) là một trong những kỹ thuật được nhiều nhóm nghiên cứu sử dụng để phân tích kết cấu hình ảnh [22]. Bên cạnh đó, biểu đồ định hướng Gradient HOG (Histograms of Oriented Gradients) và mẫu nhị phân cục bộ LBP (Local Binary Patterns) [23] là hai bộ mô tả kết cấu với số chiều nhỏ được sử dụng phổ biến trong trích xuất đặc trưng. Các phương pháp tiếp cận cấu trúc nhằm xác định kết cấu nguyên thủy và các quy tắc sắp xếp như phát hiện biên với LoG (Laplacian of Gaussian) hay DoG (Difference of Gaussian) [24],... nhằm phân đoạn hình ảnh. Các phương pháp tiếp cận theo cấu trúc thường được sử dụng cho các kết cấu thông thường vì tính đều đặn, lặp lại định kỳ trong kết cấu với một số quy tắc sắp xếp, thường độc lập với các phép biến đổi hình học như phép tịnh tiến, phép quay và phép chia tỷ lệ. Tuy nhiên, các phương pháp này thường dùng cho mục đích tổng hợp hơn là mục đích phân tích, do chúng không được sử dụng cho các kết cấu có mức độ ngẫu nhiên cao. Các phương pháp biến đổi thể hiện một hình ảnh trong một không gian mà hệ tọa độ của nó liên quan chặt chẽ đến các đặc trưng của kết cấu nhằm phân đoạn hình ảnh [21] như: biến đổi Fourier phân tích nội dung của kết cấu theo miền tần số, phép lọc Gabor và phép biến đổi Wavelet phân tích nội dung của kết cấu cả trong miền tần số và miền không gian. Như vậy, có nhiều kỹ thuật khác nhau để trích xuất đặc trưng kết cấu cấp thấp của hình ảnh. Tuy nhiên, đặc trưng kết cấu độ nhạy với nhiễu ảnh và ngữ nghĩa phụ thuộc vào hình dạng đối tượng ảnh. Do đó, các xu hướng gần đây thường kết hợp đặc trưng kết cấu với đặc trưng hình dạng cho bài toán tìm kiếm ảnh theo ngữ nghĩa.

Đặc trưng hình dạng

Hình dạng là một đặc trưng mức thấp nhằm nhận dạng đối tượng trong hình ảnh, ổn định với những thay đổi về ánh sáng, màu sắc và kết cấu [18]. Tìm kiếm ảnh theo đặc trưng hình dạng có độ chính xác tốt với các đặc trưng nhỏ gọn, độ phức tạp tính toán thấp. Các kỹ thuật biểu diễn và mô tả hình dạng có thể được phân thành hai loại: phương pháp dựa trên đường biên và phương pháp dựa trên vùng [138].

Các kỹ thuật trích xuất đặc trưng dựa trên đường biên của hình dạng như phương pháp chia nhỏ đường biên của hình dạng thành các đoạn nhỏ và dựa trên các đặc điểm hình học của nó [24] như tạo thành chuỗi đặc trưng, dựa trên phép lọc Sobel, phát hiện cạnh Canny, hay phát hiện biên với đường cong Bezier và đường cong B-spline,... Trong phương pháp dựa trên vùng, tất cả các pixel trong một vùng hình dạng được tính toán để biểu diễn hình dạng, với các kỹ thuật [24] như: biến đổi Wavelet để phân vùng các đặc điểm tương tự nhau trong ảnh, biến đổi Hough cho từng điểm cạnh trên các hình dạng bị biến dạng và nhiễu, mô men Zernike trích xuất thông tin toàn cục của hình ảnh, mô men Pseudo-Zernike bất biến và ít nhạy cảm với nhiễu hơn mô men Zernike,… Các thực nghiệm với phương pháp trích xuất đặc trưng dựa trên vùng cho thấy sự vượt trội hơn so với phương pháp dựa trên đường biên, do nó sử dụng hiệu quả tất cả thông tin pixel trong vùng đối tượng, tuy nhiên, cũng vì thế mà phương pháp này có kích thước lớn và phức tạp về tính toán hơn.

Bảng 1.1: Bảng mô tả kí hiệu và đặc tính của các đặc trưng mức thấp (LF- Low level features) được sử dụng trong các thực nghiệm của LA

Các mô tả	Kiểu đặc trưng	Số chiều vector đặc trưng	Ghi chú
GCM	Màu sắc	81	CT1 - CT8
LBP	Kết cấu	59	CT1 - CT8
GWT	Kết cấu	120	CT1 - CT8
EDH	Hình dạng	37	CT1 - CT8
GIST	Hình dạng	512	CT1 - CT8
HOG	Hình dạng	720	CT5

Thông thường để nâng cao hiệu quả trong CBIR, gần đây các nghiên cứu đã sử dụng kết hợp các đặc trưng mức thấp về cả màu sắc, kết cấu và hình dạng để tạo thành bộ mô tả đặc trưng như: kết hợp giữa đặc trưng màu sắc dựa trên biểu đồ màu HSV và đặc trưng kết cấu được trích xuất bằng Biến đổi Wavelet rời rạc DWT (Discrete Wavelet Transform), bộ mô tả biểu đồ biên EDH (Edge Histogram Descriptor) [25]; sử dụng không gian màu RGB, GLCM để trích xuất các cạnh và góc của hình dạng đối tượng [23]; kết hợp mô men màu, Gabor Wavelet và biến đổi rời rạc Wavelet, cùng với bộ mô tả hướng màu và cạnh cho đặc trưng cấp thấp [16]… Các nghiên cứu này cho thấy hiệu quả của phương pháp tìm kiếm ảnh với bộ đặc trưng kết hợp vượt trội hơn so với các phương pháp chỉ sử dụng một loại đặc trưng.

Trong luận án này, phương pháp trích xuất và kết hợp các đặc trưng màu sắc, kết cấu và hình dạng được đề xuất với bảng mô tả các đặc trưng kết hợp, số chiều vector các đặc trưng mức thấp được sử dụng theo Bảng 1.1.

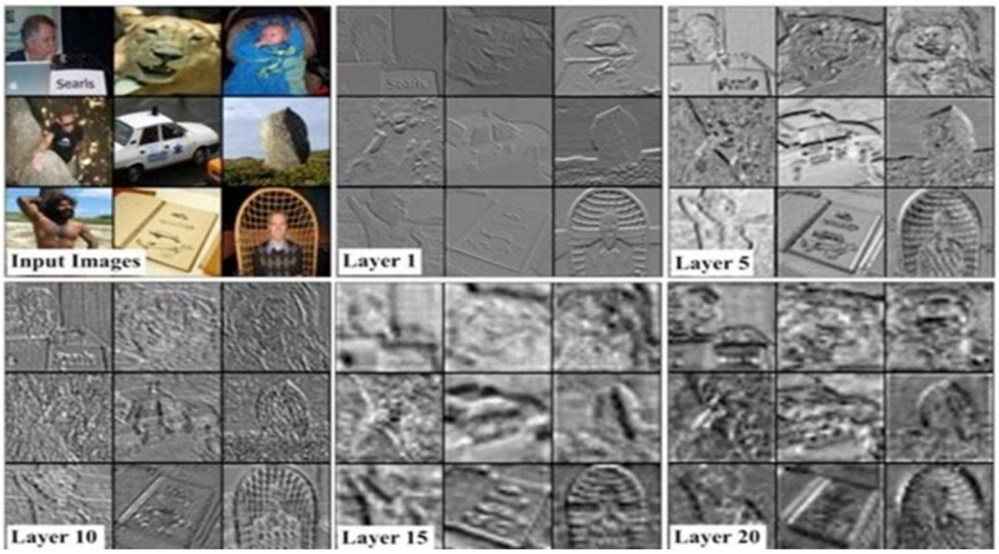
1.2.2 Đặc trưng mức cao của ảnh

Hiệu quả tra cứu ảnh sử dụng biểu diễn đặc trưng mức thấp như trên (gọi là các đặc trưng thủ công - handcraft) có điểm hạn chế bởi vì những đặc trưng

thủ công này có thể không mô tả được ngữ nghĩa của ảnh, không thể tận dụng hiệu quả các vùng nổi trội và thường bỏ qua cấu trúc không gian của hình ảnh [26].

Mạng CNN đã có nhiều tiến bộ vượt bậc trong mấy năm gần đây và đã đạt được nhiều thành tựu ấn tượng để giải quyết các bài toán thị giác máy tính [7], chẳng hạn như nhận dạng khuôn mặt, dò tìm và nhận dạng biển số xe. Như tên gọi của mạng, các mạng CNN bao gồm nhiều lớp tích chập (Convolutional layers), mỗi lớp gồm nhiều bộ lọc với cùng kích thước (thường rất nhỏ: n×n với n = {3,5,7,11}). Mỗi bộ lọc sẽ có các trọng số và được dùng chung trên mọi vị trí của ảnh đầu vào, được gọi là kỹ thuật chia sẻ trọng số (weight sharing). Các trọng số ban đầu được khởi tạo giá trị ngẫu nhiên và sẽ được cập nhật trong quá trình huấn luyện bởi các hàm mục tiêu nhằm hội tụ về trạng thái phù hợp với ngữ cảnh của một bài toán cụ thể.

Các tầng phía đầu mạng sẽ học các đặc trưng mức thấp như thông tin về hình dạng, cạnh, các điểm góc, màu sắc, hình dạng. Càng về cuối mạng, các đặc trưng càng có xu hướng giàu thông tin hơn, chứa đựng nhiều đặc trưng ngữ cảnh hơn, giúp mô hình phân biệt được các đối tượng khác nhau, minh hoạ ở hình 1.5.



Hình 1.5: Đặc trưng ảnh được trích xuất tại các lớp của mạng CNN [147]

Đầu ra của mạng CNN có thể là một hoặc nhiều vector đặc trưng được tạo ra bằng cách cho ảnh đầu vào chạy xuyên qua mạng với các trọng số đã được huấn luyện. Tùy thuộc vào từng bài toán khác nhau, chúng ta có thể sử dụng các vector đặc trưng đó theo nhiều cách khác nhau. Chẳng hạn, có thể sử dụng một bộ phân lớp cơ bản để phân lớp các vector đặc trưng, hoặc có thể gắn một tầng phân lớp (chẳng hạn dùng hàm SoftMax) vào cuối của mô hình mạng để dự đoán nhãn của các vector đặc trưng.

DML và đặc trưng Véc tơ nhúng (Embedded Vector features)

DML là việc sử dụng các kiến trúc sâu để giải quyết các hạn chế gây ra từ dữ liệu thô bằng cách có được đặc trưng nhúng tương tự thông qua học tập không gian con phi tuyến [128]. Việc sử dụng dữ liệu hình ảnh đầu vào với vector nhúng đặc trưng được trích xuất từ DML, đã được chứng minh là mang lại hiệu quả trong phân lớp và phân cụm. Nhờ khả năng phân biệt hiệu quả, DML được áp dụng trong nhiều lĩnh vực khác nhau, chẳng hạn như xác minh và nhận dạng khuôn mặt, mô hình ba chiều (3D), Xác minh chữ ký [129-131], v.v.

DML bao gồm ba phần chính là thông tin các mẫu đầu vào, cấu trúc của mô hình mạng và hàm mất mát. Lựa chọn mẫu đầu vào đóng vai trò rất quan trọng để đảm bảo thành công của DML trong phân lớp hoặc phân cụm. Mạng Siamese, Triplet và Quadruple được sử dụng phổ biến nhất để huấn luyện các mẫu trong DML, tuy nhiên, trong đó mạng Triplet đơn giản nhưng được chứng minh là đạt hiệu suất cao trong xác minh và nhận dạng khuôn mặt [130].

Véc tơ nhúng là đặc trưng được trích xuất từ các mạng học sâu đã huấn luyện sẵn. Chúng mang lại một biểu diễn cao cấp hơn so với đặc trưng mức thấp, phản ánh nhiều ý nghĩa ngữ nghĩa hơn trong hình ảnh.

Cụ thể, véc tơ nhúng được trích xuất từ lớp kết nối đầy đủ (fully connected) cuối cùng của mô hình học sâu ResNet50 (mạng phần dư) đã được huấn luyện trước. Mô hình này được xây dựng dựa trên kiến trúc một khối phần dư (residual block), giúp thông tin được truyền qua các tầng lớp sâu hơn một cách hiệu quả.

Việc trích xuất từ lớp fully connected cuối cùng có thể phản ánh được cấu trúc phức tạp của dữ liệu đầu vào thông qua quá trình biến đổi tuyến tính phi tuyến. Đặc biệt, với mô hình được huấn luyện sâu bằng phương pháp học sâu metric như Triplet loss, các véc tơ này mang tính phân biệt cao giữa các lớp khác nhau.

Véc tơ nhúng thể hiện đặc trưng của hình ảnh ở cấp độ cao hơn so với các đặc trưng mức thấp truyền thống. Tuy nhiên, chúng cũng gặp một số hạn chế như phụ thuộc vào khả năng huấn luyện của mô hình sâu và có thể bị ảnh hưởng bởi sự khác biệt giữa dữ liệu huấn luyện và dữ liệu thực. Điều này đòi hỏi phải kết hợp đặc trưng nhằm phát huy ưu điểm của các loại đặc trưng.

Trong luận án này, tác giả chọn mạng Triplet làm mô hình tiền huấn luyện để trích xuất các đặc trưng vector nhúng. Mạng Triplet sử dụng độ đo khoảng

cách Euclidean và bộ ba mẫu có chứa neo, điểm dương và điểm âm với hàm mất mát bộ ba (triplet loss) được định nghĩa như sau:

$$L_{Triplet} = max(0, \|G_W(X) - G_W(X^p)\|_2 - \|G_W(X) - G_W(X^n)\|_2 + \alpha) \quad (1.1)$$

Trong đó X_1, X_2 là một cặp đầu vào trong tập huấn luyện, X^n là mẫu tương tự, X^p là mẫu không tượng tự; $G_W(X_1)$, $G_W(X_2)$ là đại diện mới của một cặp mẫu đầu vào, α là lề cố định. D_W được sử dụng để tính khoảng cách giữa hai đầu vào trong hàm mất mát.

$$D_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\|_2 \quad (1.2)$$

Mục đích chính của việc học dựa trên Triplet loss là tìm giá trị sao cho khoảng cách từ điểm neo đến điểm dương gần với neo hơn điểm âm, từ đó cải thiện khả năng phân biệt của các mẫu. Dữ liệu thô ban đầu được chuyển đổi qua các mạng bộ ba và sau đó tính toán độ tương tự của khoảng cách để có được các đặc trưng vector nhúng cuối cùng chứa ngữ nghĩa và độ phân biệt cao hơn.

Ngoài ra, mạng Siamese [133] cũng được lựa chọn làm mô hình tiền huấn luyện để trích xuất các véc tơ nhúng và trích xuất đặc trưng nhúng. Mạng nơ-ron SigNet [133] dạng Siamese là một lớp kiến trúc mạng thường chứa hai phần mạng con đồng nhất. Các mạng nơ-ron tích chập (CNNs) song sinh có cùng cấu hình với các tham số và trọng số chia sẻ giống nhau. Việc cập nhật tham số được thực hiện đối xứng trên cả hai phần mạng con. Những phần mạng con này được kết hợp thông qua một hàm mất mát ở phía trên, tính toán một độ đo tương tự liên quan đến khoảng cách Euclid giữa biểu diễn đặc trưng ở mỗi bên của mạng Siamese. Một hàm mất mát phổ biến được sử dụng chủ yếu trong mạng Siamese là hàm mất mát tương phản, được định nghĩa như sau:

$$L_{Contrastive} = \alpha(1-y)\left\|f_{w_1}(s_1) - f_{w_2}(s_2)\right\|_{L_2}^2 + \beta \max\left(0, m - \left\|f_{w_1}(s_1) - f_{w_2}(s_2)\right\|_{L_2}^2\right) \quad (1.3)$$

trong đó s_1 và s_2 là hai mẫu (ở đây là hình ảnh chữ ký), y là hàm chỉ số nhị phân đánh dấu xem hai mẫu có thuộc cùng một lớp hay không, α và β là hai hằng số và m là giới hạn bằng 1, f là một hàm nhúng chuyển đổi hình ảnh chữ ký thành không gian véc tơ thực thông qua CNN, và w_1, w_2 là các trọng số học được cho một lớp cụ thể của mạng cơ sở. Khác với các phương pháp tiếp cận thông thường gán nhãn tương tự nhị phân cho các cặp, mạng Siamese nhằm đưa các vectơ đặc trưng đầu ra gần nhau cho các cặp đầu vào được gán nhãn là tương tự, và đẩy các vectơ đặc trưng xa nhau nếu các cặp đầu vào không tương tự. Mỗi nhánh của mạng Siamese có thể được xem như một hàm nhúng hình ảnh đầu vào vào một không gian. Do hàm mất mát được chọn (công thức 1.3), không gian này sẽ có tính chất rằng hình ảnh của cùng một lớp (chữ ký thật của một người viết cụ thể) sẽ gần nhau hơn so với hình ảnh của các lớp khác nhau (chữ ký giả mạo hoặc chữ ký của các người viết khác nhau). Cả hai nhánh được kết hợp với nhau thông qua một lớp tính toán khoảng cách Euclid giữa hai điểm trong không gian nhúng. Sau đó, để quyết định hai hình ảnh có thuộc cùng lớp tương tự (thật, thật) hay lớp không tương tự (thật, giả) cần xác định một giá trị ngưỡng dựa trên giá trị khoảng cách.

Các lớp tích chập đầu tiên lọc hình ảnh chữ ký đầu vào kích thước 155 × 220 bằng 96 kernel kích thước 11×11 với bước đi là 1 pixel. Lớp tích chập thứ hai nhận đầu vào từ kết quả (đã được chuẩn hóa và thu gọn) của lớp tích chập đầu tiên và lọc nó bằng 256 kernel kích thước 5 × 5. Lớp tích chập thứ ba và thứ tư được kết nối với nhau mà không có sự can thiệp của lớp thu gọn hoặc chuẩn hóa. Lớp thứ ba có 384 kernel kích thước 3 × 3 được kết nối với đầu ra (đã chuẩn hóa, thu gọn và thực hiện dropout) của lớp tích chập thứ hai được đưa vào lớp tích chập thứ tư. Lớp tích chập thứ tư có 256 kernel kích thước 3×3. Điều này dẫn đến mạng nơ-ron học các đặc trưng cấp thấp ít hơn cho các vùng tiếp nhận nhỏ hơn và nhiều đặc trưng cho các đặc trưng cấp cao hoặc trừu

tượng hơn. Lớp kết nối đầy đủ đầu tiên có 1024 neuron, trong khi lớp kết nối đầy đủ thứ hai có 128 neuron. Điều này cho thấy rằng véc tơ đặc trưng đã học cuối cùng của SigNet có kích thước bằng 128.

Bảng 1.2: Bảng tổng quan về các mạng CNN

Layer	Size	Parameters
Convolution	96×11×11	stride = 1
Local Response Norm	-	$\alpha = 10^{-4}$; $\beta = 0.75$, $k = 2$; $n = 5$
Pooling	96×3×3	stride = 2
Convolution	256×5×5	stride = 1, pad = 2
Local Response Norm	-	$\alpha = 10^{-4}$; $\beta = 0.75$, $k = 2$; $n = 5$
Pooling + Dropout	256×3×3	stride = 2; p = 0:3
Convolution	384×3×3	stride = 1; pad = 1
Convolution	256×3×3	stride = 1; pad = 1
Pooling + Dropout	256×3×3	stride = 2; p = 0:3
Fully Connected + Dropout	1024	p = 0:5
Fully Connected	128	

Đặc trưng CNN (CNN features)

Trong thời gian gần đây, các hệ thống CBIR đã áp dụng các đặc trưng được trích xuất từ mạng học sâu để cải thiện khả năng tra cứu ảnh [5]. Việc này đã góp phần tăng tính chính xác và độ tin cậy trong quá trình tìm kiếm các hình ảnh tương tự từ cơ sở dữ liệu. Trong cách tiếp cận học sâu, một mô hình có thể xử lý dữ liệu ảnh gốc và tự khám phá ra đặc trưng tốt thông qua quá trình học. Trong [27], mô hình mạng nơ ron tích chập được sử dụng để trích rút đặc trưng cho mỗi ảnh, giúp cải thiện việc tra cứu ảnh tương tự với ảnh truy vấn tốt hơn. Mô hình bao gồm các lớp tích chập (convolutional layer), các lớp gộp (pooling)

và lớp kết nối đầy đủ (fully connected layer). Các lớp phía trước thường là các lớp tích chập kết hợp với các hàm kích hoạt phi tuyến và lớp pooling (được gọi chung là ConvNet), do vậy, đầu ra ở lớp gần cuối cùng trước khi chuyển qua lớp kết nối đầy đủ có thể được coi là véctơ đặc trưng hữu ích. Lớp cuối cùng là một mạng nơ ron kết nối đầy đủ và thường là một hàm Softmax. Desai và cộng sự [28] đã đề xuất một phương pháp CBIR dựa trên VGG16 để trích rút đặc trưng kết hợp phân lớp SVM, phương pháp này được thực nghiệm trên tập dữ liệu Corel10K và chỉ ra độ chính xác trong tra cứu tốt hơn. Trong các nghiên cứu [29, 30] đã đánh giá toàn diện về các mạng học sâu được sử dụng trong CBIR, công trình đã đề xuất sử dụng các mạng học sâu như: MobileNet, Xception, DenseNet, InceptionResNet, EfficientNet-B1 trên các tập dữ liệu Correl, Inria Holidays cho độ chính xác vượt trội.

Với việc sử dụng các mạng CNN hiện đại, chúng ta có thể trích rút các đặc trưng phức tạp và có ý nghĩa trong một tập dữ liệu lớn. Tuy nhiên, một vấn đề với việc sử dụng các đặc trưng CNN là chúng thường rất phức tạp và có số chiều lớn điều này có thể gây ra vấn đề về tốc độ xử lý và bộ nhớ khi sử dụng các đặc trưng này trong các hệ thống CBIR. Do đó, một số nghiên cứu đã tinh chỉnh các mạng CNN để giảm số lượng tham số và kích thước của các đặc trưng, hoặc sử dụng một số lớp tiềm ẩn để trích xuất các đặc trưng mức cao hơn mà vẫn giảm được số lượng tham số và kích thước của các đặc trưng.

Trong luận án này, tác giả sử dụng mạng EfficientNet [122] là một dòng mô hình mạng nơ-ron sâu tiên tiến, được phát triển để đạt được hiệu suất và độ chính xác tối ưu trong lĩnh vực học máy.

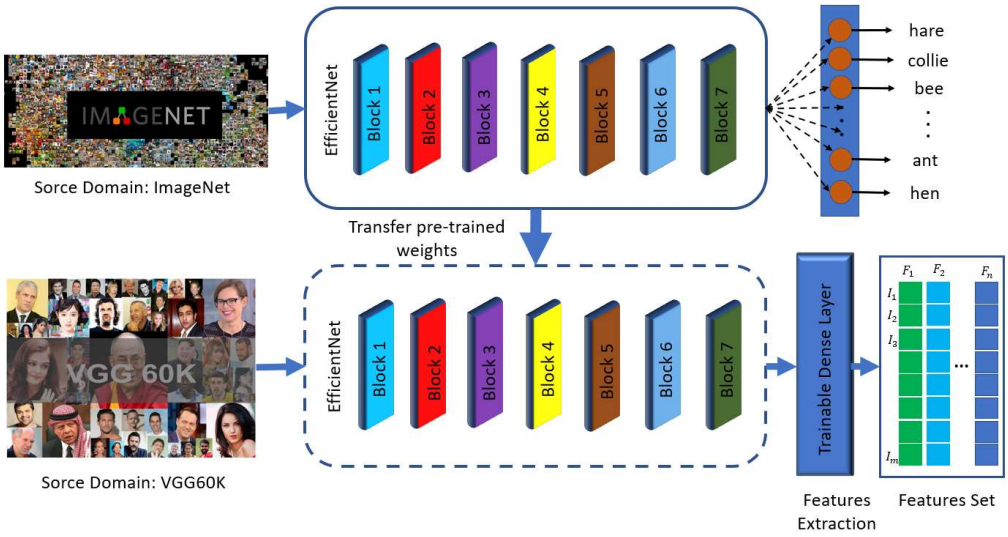
Bảng 1.3: Kiến trúc mạng EfficientNet-B0

Stage	Operator	Resolution	Chanels	Layers
i	$\widehat{F_i}$	$\widehat{H_i}\ x\ \widehat{W_i}$	$\widehat{C_i}$	$\widehat{L_i}$
1	Conv3x3	224x224	32	1
2	MBcon1, k3x3	112x112	16	1
3	MBcon6, k3x3	112x112	24	2
4	MBcon6, k5x5	56x56	40	2
5	MBcon6, k3x3	28x28	80	3
6	MBcon6, k5x5	14x14	112	3
7	MBcon6, k5x5	14x14	192	4
8	MBcon6, k3x3	7x7	320	1
9	Conv1x1 & Pooling & FC	7x7	1280	1

Các tác giả của EfficientNet đã đề xuất một phương pháp mới để mở rộng chiều rộng, chiều sâu và độ phân giải của mạng thông qua việc sử dụng hệ số kết hợp. Sự mở rộng này được thực hiện thông qua Neural Architecture Search, giúp tạo ra các phiên bản như EfficientNet-B0 đến EfficientNet-B7, với kích thước mạng giảm và hiệu suất tăng đáng kể. EfficientNet đạt đến độ chính xác tốt nhất trên nhiều tập dữ liệu và là một lựa chọn mạnh mẽ cho các ứng dụng thị giác máy tính và học sâu.

Như đã đề cập ở trên, luận án sử dụng EfficientNet như một trích xuất đặc trưng cho hệ thống CBIR. Mô hình này đã được huấn luyện trên ImageNet với 14 triệu hình ảnh được gán nhãn vào 21K nhóm và tiến hành cắt giảm các lớp cuối cùng của mạng này để có được một mô hình mạng để trích xuất vector

đặc trưng. Sau khi bộ dữ liệu hình ảnh đi qua mạng EfficientNet này thu được tập vector đặc trưng với kích thước 1280 chiều.



Hình 1.6: Mô hình trích rút đặc trưng ảnh bằng mô hình học sâu [122]

Hình 1.6 là thuật toán biểu diễn đặc trưng ảnh được trích rút dựa vào tiền huấn luyện mạng học sâu CNN thu được tập đặc trưng CNN. Với đầu vào của thuật toán là CSDL ảnh và mô hình tiền huấn luyện các mạng học sâu CNN trên bộ dữ liệu ImageNet được thực hiện các qua các bước như sau:

Thuật toán 1.1: Trích rút đặc trưng ảnh với mạng CNN tiền huấn luyện

Input:

- Tập ảnh cần trích rút đặc trưng: $I=(I_1, I_2, \ldots, I_n)$

- Mô hình tiền huấn luyện Ω = *(VGG16, Resnet50, EfficientNet….)*

Ouput: Vector đặc trưng biểu diễn ảnh $HF=(HF_1, HF_2, \ldots, HF_n)$

1. Model←LoadModel(**Ω**)
2. $HF \leftarrow \phi$
3. *for i=1,…, n do*
   3.1 HF_i←ExtractFeature$(I_i, Model)$;

3.2 $HF = HF \cup HF_i$

4. Return HF

Các phương pháp tra cứu ảnh kể trên có thể nâng cao độ chính xác khi sử dụng cách tiếp cận học sâu, tuy nhiên chúng khá tốn thời gian để xử lý do số chiều của véc tơ đặc trưng thu được khá lớn và vẫn gặp phải vấn đề khoảng cách giữa đặc trưng mức thấp với cảm nhận trực quan của con người khi mô tả nội dung ảnh.

1.3. Chuẩn hoá các vector đặc trưng biểu diễn ảnh

Mục đích của việc chuẩn hóa các thành phần của vector đặc trưng là đảm bảo mỗi thành phần nhận trong số cân bằng nhau trong việc xác định độ tương tự giữa hai vector đặc trưng.

Tiếp nối quá trình biểu diễn và trích chọn các đặc trưng của ảnh, các phương pháp chuẩn hóa các đặc trưng trong CBIR đã cải thiện đáng kể hiệu năng cho CBIR qua các nghiên cứu Ciocca cùng các cộng sự [13], Rui cùng các cộng sự [71]. Trong CBIR thường sử dụng 2 phép chuẩn hóa: (1) chuẩn hóa *min-max* và (2) chuẩn hóa *Gauss.*

1.3.1. Chuẩn hóa Min-max

Chuẩn hóa min-max cho phép chuyển dữ liệu về phạm vi [0,1] như sau:

$$\min - \max: \mathrm{f_i} = \{\mathrm{f_{i,j}}\} \mapsto \mathrm{f_i'}\{\mathrm{f_{i,j}'}\}, \mathrm{f_{i,j}'} = \frac{\mathrm{f_{i,j}} - \min\limits_{\overline{\mathrm{E_i}}}\{\mathrm{E_{i,j}}\}}{\max\limits_{\overline{\mathrm{E_i}}}\{\mathrm{E_{i,j}}\} - \min\limits_{\overline{\mathrm{E_i}}}\{\mathrm{E_{i,j}}\}}, \forall \mathrm{j} = \overline{1, \dim(\mathrm{f_i})} \quad (1.4)$$

Trong đó $\min\limits_{\overline{E_i}}\{E_{i,j}\}$và $\max\limits_{\overline{E_i}}\{E_{i,j}\}$là các giá trị nhỏ nhất và lớn nhất của *chuỗi thành phần j theo bộ đặc trưng* $\overline{E_i}$*của tất cả ảnh trong dữ liệu*. Với phép chuẩn hóa này thì các thông tin hữu ích bị dịch chuyển vào một phạm vi rất hẹp trong [0,1] nếu giá trị max lớn.

Với phép chuẩn hóa min-max được sử dụng trong các nghiên cứu Ciocca

cùng các cộng sự [13] , Ortega cùng các cộng sự [62], nhược điểm của phép chuẩn hóa này là không hiệu quả do hầu hết các thông tin hữu ích bị dịch chuyển vào một phạm vi rất hẹp do vậy làm mất đi các đặc trưng rải rác.

1.3.2. Chuẩn hóa Gauss (chuẩn hóa 3σ)

Chuẩn hóa Gaussian (hay gọi là chuẩn hóa 3σ), một phép chuẩn hóa khá hiệu quả được sử dụng nhiều trong CBIR của Vũ Văn Hiệu cùng cộng sự [2], Rui cùng các cộng sự [71]. Chuẩn hóa 3σ sử dụng công thức tính:

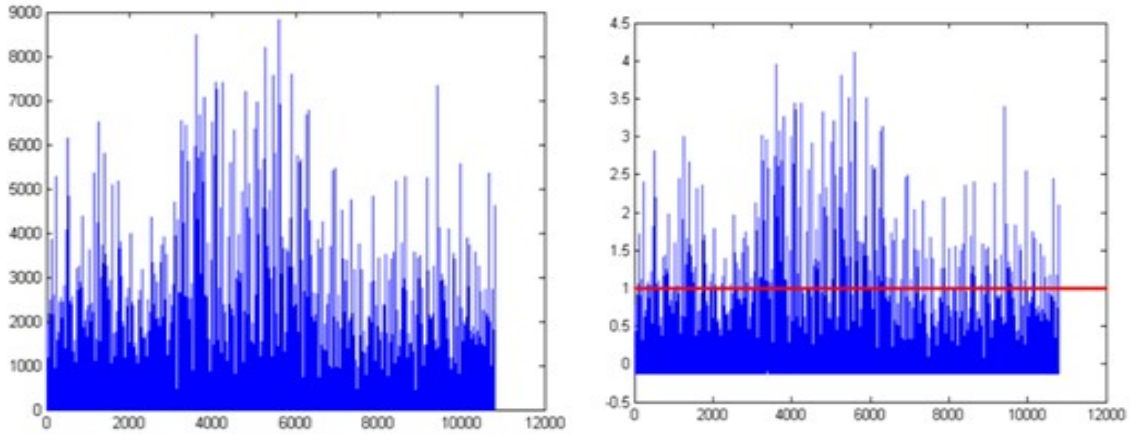
$$3\sigma: f_i = \{f_{i,j}\} \mapsto f_i'\{f_{i,j}'\}, f_{i,j}' = \frac{f_{i,j} - m_j}{3\sigma_j} \forall j = \overline{1, \dim(f_i)} \quad (1.5)$$

Trong các nghiên cứu [14], Rajakumar cùng các cộng sự [70], Rui cùng các cộng sự [71], phép chuẩn hóa 3σ được sử dụng cho các đặc trưng mức thấp (màu, kết cấu, hình dạng). Từ thực nghiệm và kết quả của các chuẩn hóa, ta thấy một số hạn chế khi chuẩn hóa theo min-max và chuẩn hóa theo 3σ là:

- Chuẩn hóa min-max làm cho hầu hết các thông tin hữu ích bị chuyển vào một phạm vi rất hẹp trong [0,1] nếu giá trị max lớn.

- Chuẩn hóa 3σ rải đều trong [-1,1] yêu cầu dữ liệu là một chuỗi Gauss.

Trong lược đồ dữ liệu Hình 1.7 các thành phần của chuỗi đặc trưng thường có không ít hơn một đỉnh, tức là giả định phân bố chuẩn áp đặt là không hợp lý. Do đó khi chuẩn hóa theo 3σ, dữ liệu sau khi chuẩn hóa có khá nhiều thành phần rơi ra ngoài đoạn [-1,1]. Vì vậy sử dụng chuẩn hoá cho dữ liệu đặc trưng kết hợp là chưa đạt được mục tiêu của bước chuẩn hoá.



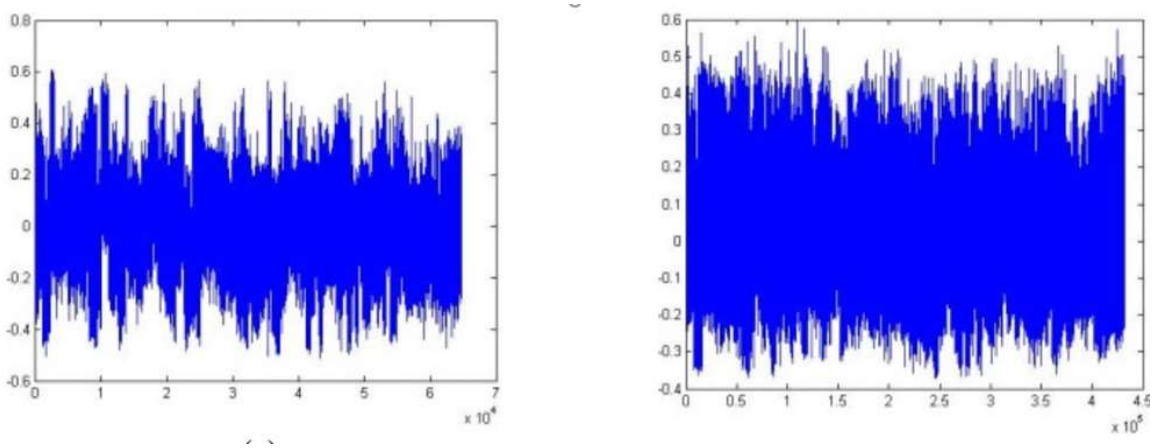
(a) (b)

Hình 1.7: Lược đồ đặc trưng

(a) Lược đồ HSV, (b) Lược đồ HSV chuẩn hóa theo quy luật 3σ

1.3.3. Chuẩn hoá đặc trưng sử dụng phân cụm mờ c-means (Fuzzy c-mean clustering - FCM)

Nhận thấy các chuẩn hóa đặc trưng theo min-max hay chuẩn hóa đặc trưng theo 3σ có nhiều hạn chế (min-max trải ra trong đoạn [0,1]; 3σ trải ra rộng hơn [-1,1] nhưng đầu vào yêu cầu là chuỗi Gaussian) nên các tác giả đã đề xuất phương pháp mới là chuẩn hóa đặc trưng dựa vào phân cụm mờ c-mean (FCM) Vũ Văn Hiệu cùng cộng sự [2]. Việc chuẩn hoá được thực hiện theo luật $3\sigma-FCM$ xem như một mở rộng của chuẩn hóa theo luật 3σ. Phép chuẩn hóa này không yêu cầu đầu vào là một chuỗi Gaussian và khi áp dụng phép chuẩn hóa 3σ-FCM đem lại hiệu quả khá cao trong CBIR. Tuy nhiên 3σ-FCM vẫn còn điểm hạn chế, đó là miền giá trị của các thành phần vector sau khi biến đổi vẫn có thể hẹp hơn đáng kể trong đoạn [-1,1] như Hình 1.8a, 1.8b.



(a) (b)

Hình 1.8: Kết quả chuẩn hóa theo 3σ-FCM.

(a) Gabor Wavelet Texture, (b) GIST.

Sau khi phân cụm FCM tập vector đặc trưng thành cụm, việc chuẩn hoá được thực hiện xem như một mở rộng của chuẩn hóa theo luật 3σ.

Phép chuẩn hóa 3σ-FCM:

Cho $x = \{x[j]\}_{j=1}^{M_t}, \tilde{x} = \{\tilde{x}[j]\}_{j=1}^{M_t}, \forall 1 \le j \le M_t$, là vector dữ liệu đầu vào theo đặc trưng mức thấp t, vector chuẩn hóa x_{norrm} của 3σ-FCM xác định như sau:

$$\tilde{x}[j] \stackrel{def}{=} \frac{\min\limits_{1\le c\le C}\left\{\frac{x[j]-v_{t,c}[j]}{3\sigma_{t,c,j}}\right\}+\max\limits_{1\le c\le C}\left\{\frac{x[j]-v_{t,c}[j]}{3\sigma_{t,c,j}}\right\}}{C+1} \quad (1.6)$$

ở đây $\sigma_{t,c,j}$ là độ lệch chuẩn của cụm thứ c theo thành phần vector thứ t.

1.4 Độ đo khoảng cách, độ đo tương tự ảnh và học các độ đo

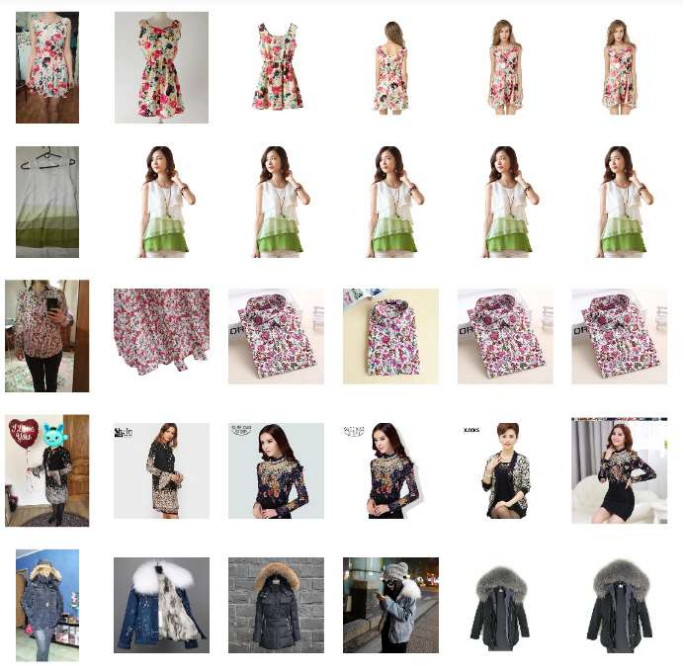
1.4.1 Độ đo khoảng cách, độ đo tương tự ảnh

Độ đo tương tự là một trong những phương pháp tốt để máy tính phân biệt được các hình ảnh qua nội dung của chúng. Thông thường hệ thống tra cứu ảnh sẽ truy vấn hình ảnh bằng phương pháp đo tương tự dựa trên các chức năng, việc xác định nó có thể dưới nhiều hình thức như phát hiện biên, màu sắc, vị trí

điểm ảnh... các phương pháp như histogram, màu sắc và phân tích histogram dòng cột sử dụng biểu đồ để xác định độ tương tự.

Bên cạnh các độ đo tương tự, trong CBIR người ta cũng áp dụng các độ đo khoảng cách để đo mức độ phản tương tự giữa các ảnh. Với một ảnh đầu vào sử dụng độ đo khoảng cách, hệ thống xếp hạng toàn bộ CSDL ảnh theo giá trị giảm dần khoảng cách các ảnh với ảnh đầu vào. Ảnh nào có giá trị khoảng cách gần với ảnh đầu vào nhất thì sẽ được sắp xếp theo thứ tự.

Các độ đo tương tự và độ đo khoảng cách được ứng trong rất nhiều lĩnh vực như xử lý ảnh và nhận dạng mẫu, nhận dạng chữ viết tay, trong y học giúp bác sĩ phát hiện các mô bệnh để tìm ra các tế bào ung thư (sử dụng công cụ tự phát huỳnh quang),…



Hình 1.9: Một ứng dụng tìm kiếm shop bán quần áo theo ảnh có sẵn. [148]

Một metric d trong Toán học (độ đo khoảng cách) trên tập hợp điểm *X,* được định nghĩa như sau: $d: X \times X \to \mathbb{R}$, thoản mãn các điều kiện sau:

$$\begin{cases} d(x,y) \geq 0 \\ d(x,y) = 0 \Leftrightarrow x \equiv y \\ d(x,y) = d(y,x) \\ d(x,z) \leq d(x,y) + d(y,z) \end{cases}$$

Các độ đo khoảng cách có thể được sử dụng cho đối sánh tương tự trong lĩnh vực CBIR như: Euclid, Mahalanobis, Minkowski…

Khoảng cách Euclid (khoảng cách L2) và khoảng cách Manhattan (khoảng cách L1) là trường hợp đặc biệt của độ đo Minkowski. Trong CBIR, khoảng cách Euclid có trọng số đã được sử dụng cho các mô men màu trong hệ thống MARS [31], khoảng cách thống kê Mahalanobis [32] được sử dụng để đo khoảng cách giữa một vectơ đặc trưng cụ thể và một phân phối đã cho. Nói chung về mặt thực tế các độ đo khoảng cách sử dụng trong CBIR không phải là một metric theo nghĩa Toán học, các tính chất đối xứng và bất đẳng thức tam giác có thể không được thỏa mãn.

Độ đo tương tự Cosine (có giá trị đo được thuộc [-1,1]) được đánh giá là phù hợp trong học độ đo tương tự, nhưng không phù hợp để biểu diễn khoảng cách giữa hai điểm dữ liệu, nhất là trên không gian đa tạp phi tuyến [134].

Một số công thức tính độ khoảng cách hay được sử dụng trong CBIR được mô tả như bảng sau:

Bảng 1.4: Một số độ đo khoảng cách, độ đo tương tự và công thức

Độ đo	Công thức tính
Euclid	$\sqrt{(x-y)^T(x-y)}$
Mahalanobis	$\sqrt{(x-y)^T C^{-1}(x-y)}$; C là psd (ma trận xác dịnh dương)
Minkowski	$D(x,y)=(\sum_{i=1}^{n}\lvert x_i-y_i\rvert^p)^{\frac{1}{p}}$
Manhattan (Taxicab/City Block) *(Khoảng cách L_1)*	$D(x,y)=\sum_{i=1}^{k}\lvert x_i-y_i\rvert$ (k là số block)
Chebyshev	$D(x,y)=\max_i(\lvert x_i-y_i\rvert)$
Cosine	$D(X,Y)=cos\,\theta=\frac{X.Y}{\lvert\lvert X\rvert\rvert.\lvert\lvert Y\rvert\rvert}$

Nếu sử dụng độ đo tương tự để xếp hạng CSDL ảnh theo ảnh truy vấn thì các ảnh được xếp hạng theo thứ tự giảm dần của giá trị đo mức tương tự với ảnh truy vấn, ngược lại nếu sử dụng một độ đo khoảng cách (đo phản tương tự) thì các ảnh của CSDL được xếp theo thứ tự tăng dần của giá trị khoảng cách.

Từ một độ đo khoảng cách d (không nhất thiết là một metric Toán học) trên $X\ x\ X$, chúng ta có thể xác được một độ đo tương tự S trên $X\ x\ X$ (giá trị tương tự đo được thuộc (0,1]) và ngược lại:

$$S(x_1,x_2)=\frac{1}{1+d(x_1,x_2)}\quad \forall x_1,x_2\in X. \tag{1.7}$$

$$\text{Ngược lại, } d(x_1,x_2)=-1+\frac{1}{S(x_1,x_2)},\ \forall x_1,x_2\in X. \tag{1.8}$$

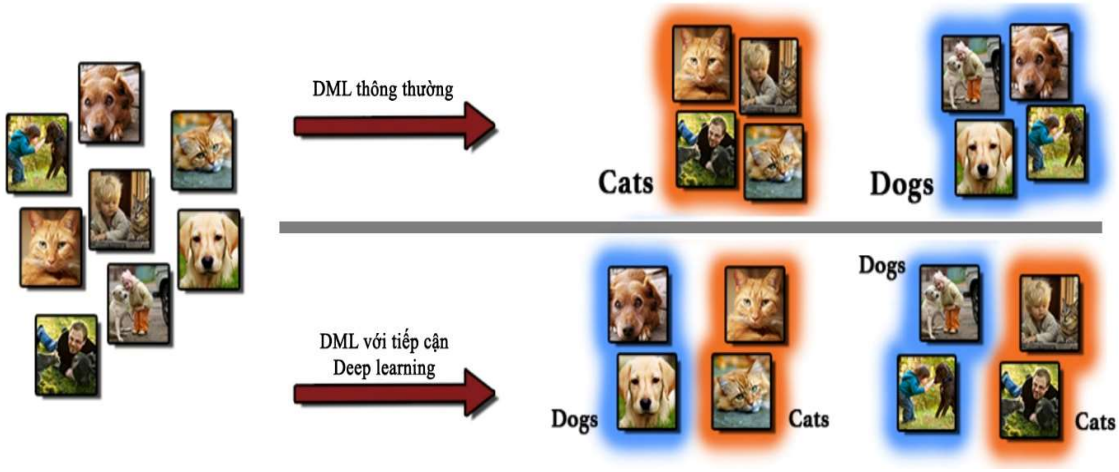
1.4.2 Học độ đo tương tự ảnh

Học độ đo tương tự là một phương pháp học máy (ML, machine learning) tập trung vào việc đánh giá mức độ tương tự hoặc khác biệt giữa các cặp dữ liệu, thay vì dự đoán nhãn hoặc tìm kiếm các cấu trúc ẩn như trong học có giám sát (supervised learning) hay học không giám sát (unsupervised learning). Mục

tiêu chính của học tương tự là học cách định lượng mức độ tương tự giữa các đối tượng, điều này có nghĩa là, với hai đối tượng bất kỳ, mô hình sẽ đánh giá xem chúng giống nhau hay khác nhau đến mức nào. Phương pháp này không yêu cầu nhãn của từng phần tử dữ liệu, nhưng cần các cặp dữ liệu được xác nhận là "tương tự" hoặc "không tương tự" để học hàm đánh giá mức độ tương tự.

Hiển nhiên, một độ đo tương tự tốt sẽ làm tăng độ chính xác của hệ thống CBIR, giúp các hệ thống này trả về các hình ảnh liên quan một cách chính xác hơn khi người dùng thực hiện truy vấn bằng hình ảnh. Một số thuật toán phổ biến trong đánh giá độ tương tự có thể kể đến như mạng Siamese, Triplet loss. Các phương pháp này đều hướng tới việc học các không gian nhúng mà trong đó các mẫu tương tự gần nhau và các mẫu khác biệt xa nhau. Siamese Networks sử dụng hai mạng con giống hệt nhau để so sánh hai điểm dữ liệu và đánh giá mức độ tương tự, trong khi Triplet Loss tạo ra các không gian đặc trưng tối ưu bằng cách huấn luyện mô hình với các bộ ba dữ liệu.

Học độ đo tương tự mở ra nhiều cơ hội tiếp cận linh hoạt cho việc xử lý các bài toán mà đánh giá mối quan hệ giữa các đối tượng là mục tiêu chính, giúp cải thiện hiệu quả và độ chính xác trong nhiều ứng dụng thực tiễn. Với khả năng đánh giá mức độ tương đồng một cách chính xác, phương pháp này không chỉ giúp tăng cường hiệu suất của các hệ thống hiện có mà còn mở ra các hướng nghiên cứu và ứng dụng mới trong nhiều lĩnh vực khác nhau.



Hình 1.10: Kết quả nhận dạng ảnh giữa sử dụng DML thông thường và DML tiếp cận học sâu [139].

1.4.2.1 Phương pháp học độ đo khoảng cách Neighbourhood Components Analysis:

Phương pháp học độ đo khoảng cách Neighbourhood Components Analysis (NCA) là một kỹ thuật học độ đo khoảng cách được đề xuất bởi Goldberger và cộng sự năm 2004 [125]. NCA không tham số áp dụng cho cả bài toán phân loại và hồi quy. Mục tiêu của NCA là tối ưu hóa trọng số của đặc trưng để tối đa hóa độ chính xác của bài toán bằng cách tìm hàm ánh xạ sao cho các điểm dữ liệu cùng nhãn gần nhau hơn các điểm khác nhãn, giúp giảm sai số phân lớp của k-NN. NCA tối ưu hóa hàm mục tiêu dựa trên sai số phân lớp dự đoán của k-NN trên tập huấn luyện bằng cách tính xác suất các láng giềng cùng nhãn với mỗi điểm dữ liệu và cực đại hóa tổng xác suất.

Cụ thể, với một tập dữ liệu gồm n mẫu đầu vào $S=\{(x_i,y_i), i=1,2,...n\}$, NCA xây dựng một mô hình phân loại/hồi quy ngẫu nhiên. Mô hình này lựa chọn ngẫu nhiên một điểm trong S làm điểm tham chiếu Ref(x) cho x. Nhãn dự đoán của x sẽ bằng nhãn của Ref(x).

Xác suất chọn $\text{Ref}(x) = x_j$ được tính dựa trên độ tương tự giữa x và x_j theo độ đo d_w, được định nghĩa như sau:

$$P\big(Ref(x) = x_j\big) = \frac{\exp\left(-d_w(x,x_j)\right)}{\sum_{k=1}^{n}\exp\left(-d_w(x,x_k)\right)} \quad (1.9)$$

Trong đó:

$d_w(x, x_j)$ là độ đo khoảng cách giữa x và x_j, dựa trên trọng số w. Độ đo này càng nhỏ thì x và x_j càng giống nhau.

$\sum_{k=1}^{n}\exp\left(-d_w(x,x_k)\right)$ là tổng của các giá trị $\exp\left(-d_w(x,x_k)\right)$ với k chạy từ 1 đến n, với n là số lượng mẫu dữ liệu.

Nhãn dự đoán của x sẽ bằng nhãn của điểm tham chiếu Ref(x).

NCA tìm trọng số w sao cho xác suất sai phân loại trung bình trên tập huấn luyện đạt giá trị nhỏ nhất, bằng cách tối ưu hóa hàm mục tiêu sử dụng phương pháp đạo hàm và bổ sung điều kiện bình thường hóa.

Ưu điểm của NCA là không giả định mô hình trước, áp dụng cho cả phân loại và hồi quy. Tuy nhiên, phương pháp cũng gặp một số hạn chế như phụ thuộc tham số, chỉ xét tính tương tự cục bộ.

1.4.2.2 Học độ đo theo nhóm phân cụm (CMML):

Học độ đo theo nhóm phân cụm (CMML - Clustered Multi-Metric Learning) là một phương pháp hiệu quả để giải quyết bài toán dữ liệu phân bố không đồng đều. Cơ sở ý tưởng của CMML bao gồm:

- Chia dữ liệu huấn luyện thành các nhóm (cluster) không giao nhau bằng k-means. Mỗi nhóm sẽ đại diện cho một vùng của không gian tính năng.

- Học một ma trận khoảng cách riêng cho mỗi nhóm dựa trên các ràng buộc triplet trong nhóm đó.

- Học thêm một ma trận khoảng cách toàn cục để khiến các ma trận cục bộ khác nhau càng gần với ma trận này...

Bên cạnh đó, một kỹ thuật điều chỉnh toàn cục được áp dụng nhằm bảo toàn các đặc tính chung của các cụm trong không gian metric đã được học.

Triplet loss được sử dụng như một hàm mất mát để học các ma trận khoảng cách trong CMML. Cụ thể:

Mỗi ràng buộc triplet bao gồm 3 điểm: điểm tâm x_i, điểm dương x_j cùng lớp với x_i và điểm âm x_l khác lớp với x_i. Mục tiêu là học ma trận khoảng cách sao cho khoảng cách từ x_i đến x_j nhỏ hơn khoảng cách từ x_i đến x_l với khoảng cách biên 1.

Đây chính là hàm mất mát Triplet loss cần tối thiểu hóa trong quá trình học CMML. Việc lựa chọn các ràng buộc triplet hợp lý giúp CMML học được các ma trận khoảng cách phân biệt tốt giữa các lớp.

Phương pháp CMML đã được chứng minh là tăng cường tính linh hoạt và hiệu quả khi áp dụng học nhiều khoảng cách được học trong các ứng dụng của học máy và khai thác dữ liệu. Kỹ thuật này đã được ứng dụng thành công trong việc xử lý dữ liệu đa dạng, và đã có những ứng dụng hiệu quả trong nhiều lĩnh vực như thị giác máy tính và xử lý ngôn ngữ tự nhiên. Phương pháp CMML đã được nhóm nghiên cứu Bac Nguyen, và cộng sự đề xuất [56].

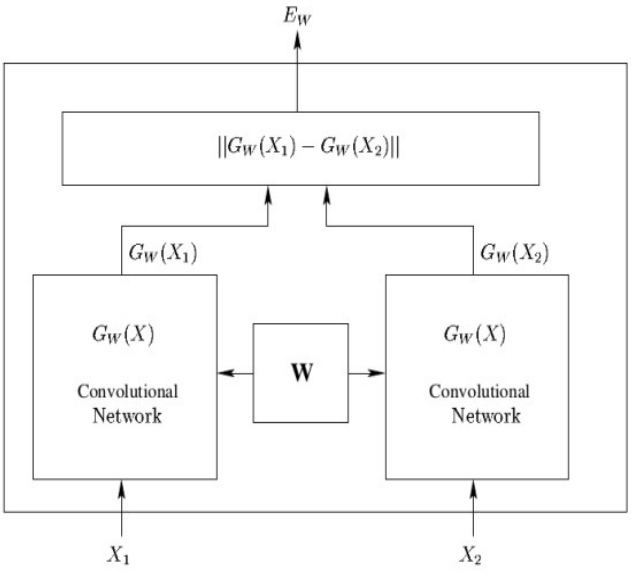
1.4.3 Tiếp cận Deep learning cho học độ đo khoảng cách

Cho một họ các hàm $G_W(X)$ được tham số hóa bởi W, cần tìm W sao cho là tập véc tơ đầu vào, khi đó:

$$D_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\|; X_1, X_2 \quad (1.10)$$

$D_W(X_1, X_2)$ nhỏ hơn ngưỡng đối với các cặp giống nhau,

$D_W(X_1, X_2)$ lớn hơn hoặc bằng ngưỡng đối với các cặp khác nhau.



Hình 1.11: Mô hình tiếp cận Deep learning cho học độ đo khoảng cách [149].

CNN có thể là một mô hình đã được huấn luyện sẵn như VGG 16, Resnet, ImageNet… hay được tự huấn luyện.

1.4.3.1 Các chiến thuật học

+ Chiến thuật Contrastive [139];

+ Chiến thuật Triplet [140];

- Chiến thuật Contrastive (mạng Siamese)

Cho $X_1, X_2 \in I$ là cặp véc tơ đầu vào, và Y là nhãn nhị phân, nếu:

Y = 0 là cặp tương tự

Y = 1 là cặp không tương tự

Khi đó hàm khoảng cách với tham số W được biểu diễn như sau

$$D_W(X_1, X_2) = \|G_W(X_1) - G_W(X_2)\|_2 \quad (1.11)$$

Hàm mất mát được đưa ra như sau

$$L(W, Y, X_1, X_2) = (1 - Y)\frac{1}{2}\left(D_W(X_1, X_2)\right)^2 + (Y)\frac{1}{2}\{max\left(0, m - D_W(X_1, X_2)\right)\}^2 \quad (1.12)$$

trong đó: m>0 là lề.

Hàm mất mát tổng thể:

$$L = \frac{1}{2N}\sum_{i=1}^{N} L\left(W, Y_i, X_{1,i}, X_{2,i}\right) \tag{1.13}$$

với N là số lượng một lô ảnh huấn luyện.

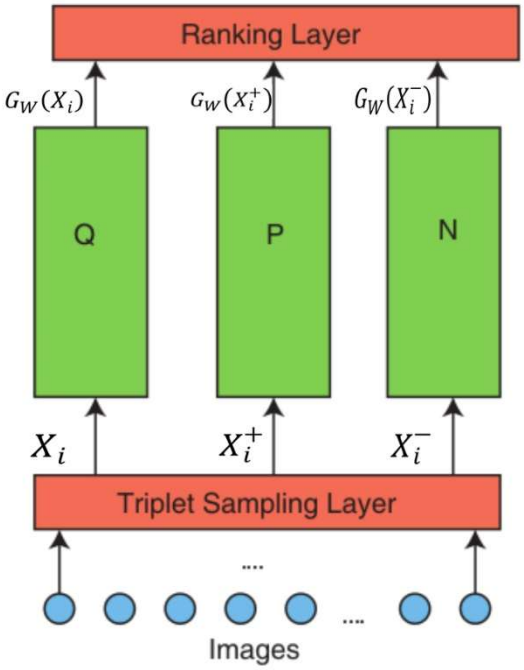
- Chiến thuật Triplet (được cải tiến từ Contrastive)

Tối ưu hóa việc nhúng và sử dụng bộ ba (ảnh neo, ảnh phù hợp và ảnh không phù hợp) để huấn luyện độ đo khoảng cách.

Với một bộ ba $(x_i,\ x_j, x_k)$ sao cho khoảng cách giữa x_i và x_j nhỏ hơn khoảng cách giữa x_i và x_k thì:

$$d_A\left(x_i; x_j\right) \leq d_A(x_i; x_k) - m \tag{1.14}$$

trong đó $m > 0$ là lề (margin).



Hình 1.12: Mô hình mạng Triplet [140]

Hàm *G(.)* được học và gán sao cho

G(.) < khoảng cách giữa các ảnh tương tự nhau.

Hàm mất mát L được định nghĩa như sau,

khi có bộ ba $T_i = (X_i, X_i^+, X_i^-)$ ta có

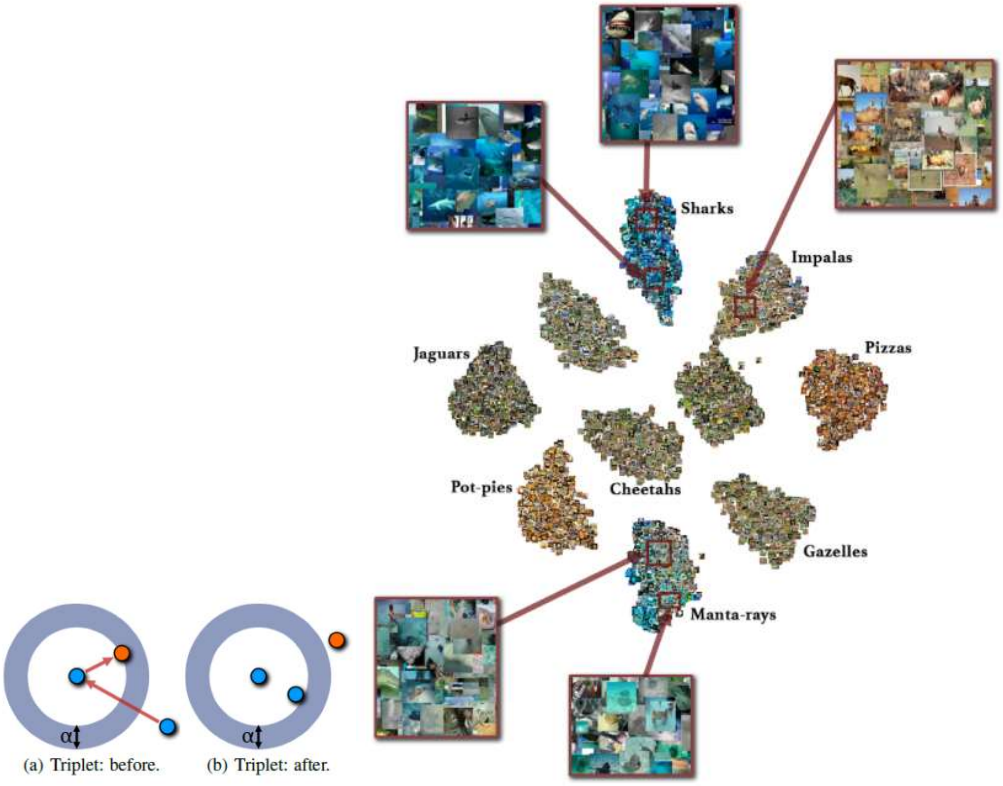
$$L(X_i, X_i^+, X_i^-) = \max\{0, m + D_W(X_i, X_i^+) - D_W(X_i, X_i^-)\} \quad (1.15)$$

trong đó: m>0 là lề (margin).

Hàm mất mát tổng thể:

$$L = \frac{1}{N}\sum_{i=1}^{N} L(X_i, X_i^+, X_i^-) \quad (1.16)$$

Thách thức: Nếu có N mẫu thì số bộ ba có thể lên tới là N^3 (Vấn đề cải tiến bộ ba triplet là vấn đề cần nghiên cứu, dù rằng đã có đưa ra [140].



Hình 1.13 Chiến thuật Triplet phân loại ảnh [140]

1.5 Xếp hạng dựa trên đồ thị với tiếp cận xếp hạng đa tạp

Các mô hình xếp hạng dựa trên đồ thị đã trở nên phổ biến trong nghiên cứu tra cứu ảnh dựa trên nội dung (CBIR) và đã thu hút sự quan tâm đáng kể từ cộng đồng học máy, thị giác máy tính và tìm kiếm thông tin trong thời gian gần đây. Chúng được áp dụng rộng rãi để mô tả mối quan hệ tương đồng/khác biệt giữa các hình ảnh, từ đó tạo ra các bảng xếp hạng hữu ích cho bài toán truy vấn.

Một số mô hình tiêu biểu có thể kể đến như: MR, AGR, SGR, EMR… Các nghiên cứu gần đây [33-35, 39, 63, CT6] đã chứng minh EMR cho kết quả xếp hạng chính xác hơn so với các phương pháp truyền thống khác. Điều này mở ra nhiều triển vọng trong ứng dụng thực tiễn.

Anchor Graph Regulrization (AGR)

Anchor Graph Regularization [38] - là mô hình xếp hạng dựa trên đồ thị sử dụng các điểm neo - anchor points để xây dựng một đồ thị thể hiện mối quan hệ giữa giữa lớp dữ liệu và lớp điểm neo [12]. Cụ thể, tính hiệu quả của AGR nằm ở hai bước:

1) Xây dựng các mối quan hệ kề giữa các điểm trong cùng một lớp dựa trên điểm neo, thay vì tính toán tất cả các mối quan hệ kề giữa các điểm dữ liệu;

2) Gán nhãn của các điểm dữ liệu từ các điểm neo dựa trên mối quan hệ kề giữa các lớp của chúng.

Scalable sub-graph Regularization (SGR)

Scalable sub-graph Regularization [39] - là mô hình xếp hạng dựa trên đồ thị nhằm giải quyết bài toán tra cứu ảnh dựa nội dung CBIR trên quy mô lớn.

Cụ thể, SGR chia dữ liệu thành các nhóm dựa trên anchor được lựa chọn bởi K-means. Sau đó, SGR xây dựng một đồ thị con dựa trên các anchor, thay vì toàn bộ dữ liệu. Đồ thị con được xây dựng thông qua ma trận trọng số Z giữa

mỗi mẫu và anchor gần nhất, cũng như ma trận kề W dựa trên Z. Tiếp theo, SGR áp dụng khung nhiễu đồ thị để tính toán số điểm xếp hạng cho các anchor. Sau đó dựa vào Z, số điểm xếp hạng của các anchor được lan truyền đến toàn bộ dữ liệu.

Nhờ xây dựng đồ thị con hiệu quả trên cơ sở anchor, SGR vừa đảm bảo khả năng mở rộng lên quy mô lớn, vừa đạt hiệu quả cao trong bài toán CBIR.

Xếp hạng đa tạp MR (Manifold Ranking)

Trong các phương pháp xếp hạng dựa trên đồ thị, xếp hạng trên cấu trúc dữ liệu đa tạp (Ranking on Data manifold) [33] là một trong những phương pháp đại diện và đã được áp dụng rộng rãi trong các ứng dụng tìm kiếm thông tin và học máy khác nhau.

Thuật toán xếp hạng đa tạp trong cấu trúc dữ liệu đa tạp thuộc mô hình học bán giám sát [9, 33], thuật toán này tập trung vào khám phá cấu trúc đa tạp nội tại[2] (Instrinsic manifold structure) của dữ liệu, từ đó đánh giá mức độ quan trọng và tương tác giữa các điểm dữ liệu trên cấu trúc đa tạp này. Bằng cách sử dụng thông tin về mối quan hệ và tương tác giữa các điểm dữ liệu, xếp hạng đa tạp (MR-Ranking Manifold) được sử dụng hiệu quả để thực hiện các tác vụ truy vấn và phân loại dữ liệu trong không gian đa tạp, nơi các điểm dữ liệu thường có sự biến đổi không đồng nhất.

Ý tưởng chính của thuật toán xếp hạng đa tạp như sau: Đầu tiên xây dựng một đồ thị có trọng số cho tất cả các điểm dữ liệu trong không gian đặc trưng (mỗi đỉnh biểu thị cho một điểm dữ liệu) bằng cách kết hợp cả dữ liệu được gán nhãn (các điểm truy vấn) và dữ liệu không được gán nhãn (cơ sở dữ liệu), MR gán cho mỗi điểm dữ liệu một điểm số xếp hạng tương đối, biểu thị mức

[2] Instrinsic manifold structure: Cấu trúc đa tạp nội tại là cấu trúc cơ bản của dữ liệu mà không phụ thuộc vào cách dữ liệu được nhúng (embedded) vào không gian Euclide hoặc bất kỳ không gian ngoại vi nào khác. Cấu trúc này phản ánh các mối quan hệ cơ bản và các đặc điểm của dữ liệu mà không bị biến dạng bởi các biến đổi không gian hoặc chiều.

độ liên quan đối với điểm truy vấn, quá trình xếp hạng lặp đi lặp lại cho đến khi hội tụ tới một tình trạng ổn định toàn cục. Các điểm chính thức được xếp hạng đại diện cho việc giống nhau giữa điểm dữ liệu và điểm truy vấn. Các điểm dữ liệu tương tự như các điểm truy vấn là những điểm xếp hạng lớn nhất.

Việc xây dựng đồ thị biểu diễn các điểm trong cơ sở dữ liệu (CSDL) theo thuật toán xếp hạng đa tạp (MR-Manifold Ranking) đã được đề xuất trong các nghiên cứu [9,33-35]. Mục tiêu chính của phương pháp MR dựa trên cách tiếp cận đồ thị là sử dụng phương pháp xác định trọng số của mỗi điểm dữ liệu so với các điểm dữ liệu khác dựa trên các thông tin toàn cục và cục bộ biểu diễn bên trong đồ thị.

Xét tập dữ liệu $X = \{x_1, x_2 \ldots\ldots, x_n\} \subset \mathbb{R}^m$, m là số chiều, n là số điểm dữ liệu; $x_q \in X$ là điểm truy vấn (hay điểm dữ liệu được gán nhãn), các điểm còn lại được xếp hạng theo mức độ liên quan của chúng đối với các điểm truy vấn.

$d: X \times X \to R$ là một độ đo khoảng cách giữa 2 điểm dữ liệu x_i và x_j (có thể dùng các độ đo Euclide, Manhattan,…), ký hiệu $d(x_i, x_j)$.

$r: X \to R$ là hàm xếp hạng, trong đó mỗi điểm x_i được gán một giá trị xếp hạng r_i. Ta có thể xem r như một vector: $r = [r_1, \ldots, r_n]^T$.

Khởi tạo $y = [y_1, \ldots, y_n]^T$, với $y_i=1$ nếu x_i là một truy vấn, ngược lại $y_i=0$.

Thuật toán 1.2: Thuật toán xếp hạng đa tạp cơ bản (MR)

Input:

- Tập dữ liệu $X = \{x_1, x_2 \ldots\ldots, x_n\} \subset \mathbb{R}^m$.

- Truy vấn q (được biểu diễn như một điểm dữ liệu)

- Số láng giềng k

Output: Vector xếp hạng $r^* \in \mathbb{R}^n$, trong đó r_i biểu thị mức độ liên quan của x_i đến truy vấn q.

Bước 0: Xây dựng đồ thị *K-NN* trên tập dữ liệu *X*.

Bước 1: Tính ma trận $W=[w_{ij}]_{n\times n}$ là ma trận kề với các phần tử w_{ij} là trọng số của các điểm dữ liệu x_i, x_j được tính theo công thức (1.1):

$$w_{ij}=\begin{cases}\exp[-d^2(x_i,x_j)/2\sigma^2] & \text{nếu } x_i \in k\text{-}NN(x_i)\\ 0 & \text{Các trường hợp khác}\end{cases} \quad (1.17)$$

Bước 2: Chuẩn hóa ma trận Laplace theo công thức $S = D^{-1/2}WD^{-1/2}$

(với $D_{ii} = \sum_{j=1}^{n} w_{ij}$ là tổng hàng thứ *i* của ma trận *W*)

Bước 3: Lặp giải phương trình: $r(t+1) = \alpha S r(t) + (1-\alpha)y$ (1.18)

Với tham số $\alpha \in (0,1)$ đến khi hội tụ.

Return.

Dạng công thức đóng để tính r^*: $r^* = (I_n - \alpha S)^{-1}y$ (với I_n là ma trận đơn vị).

Theo [125, 126] thì công thức (1.18) hội tụ và $r^* = (1-\alpha)(I-\alpha S)^{-1}y$ - với *I* là ma trận đơn vị cỡ $n\times n$. (công thức được chứng minh ở PL2)

Trong thực tế, hệ số $\beta = (1-\alpha)$ không ảnh hưởng đến các điểm số xếp hạng, do vậy: $r^* \approx (I-\alpha S)^{-1}y$ (1.19)

Để tính toán hiệu quả hàm xếp hạng (1.19), trong [125, 126] đã đề xuất hàm chi phí xác định độ đo xếp hạng được chuẩn hóa bởi phương trình:

$$O(r)=\frac{1}{2}\left(\sum_{i,j=1}^{n} w_{ij}\left\|\frac{1}{\sqrt{D_{ii}}}r_i-\frac{1}{\sqrt{D_{jj}}}r_j\right\|^2+\mu\sum_{i=1}^{n}\|r_i-y_i\|^2\right) \quad (1.20)$$

Hàm chi phí *O(r)* bao gồm hai thành phần chính:

+ Phần đầu tiên liên quan đến cấu trúc đa tạp nội tại của tập dữ liệu, nghĩa là mô hình hóa mối quan hệ giữa các điểm dữ liệu dựa trên cấu trúc không gian nội tại của chúng.

+ Phần thứ hai đảm bảo rằng giá trị xếp hạng mới r_i không chênh lệch quá xa so với giá trị xếp hạng ban đầu y_i.

Trong đó:

+ r là vector xếp hạng mà ta muốn tìm.

+ w_{ij} là một phần tử của ma trận trọng số W, là ma trận kề thể hiện mức độ tương quan hoặc khoảng cách giữa điểm dữ liệu thứ i và j.

+ D là ma trận đường chéo mà mỗi phần tử đường chéo D_{ii} là tổng của hàng thứ i của ma trận W.

+ μ là một tham số điều chỉnh, cân bằng giữa hai phía của hàm chi phí.

+ y_i là giá trị xếp hạng ban đầu của điểm dữ liệu thứ i.

MR (Manifold Ranking) đã được sử dụng rộng rãi trong nhiều ứng dụng, tuy nhiên để xử lý cơ sở dữ liệu quy mô lớn đã có những hạn chế xảy ra:

1- Việc xây dựng đồ thị của MR bằng đồ thị *k-NN* là không khả thi đối với dữ liệu lớn, vì chi phí xây dựng đồ thị *k-NN* là $O(n^2 \log k)$.

2- Việc xếp hạng đa tạp cũng như nhiều thuật toán dựa trên đồ thị khác trực tiếp sử dụng ma trận kề W trong việc tính toán. Chi phí lưu trữ của một ma trận thưa W là $O(kn)$.

3- Việc xếp hạng đa tạp có độ phức tạp tính toán lớn vì khai thác ma trận nghịch đảo trong phương trình (1.19).

4- Khi mở rộng CSDL thì phải tính toán và xây dựng lại toàn bộ đồ thị.

Như vậy, chúng ta cần phải tìm cách để xây dựng một đồ thị sao cho thời gian xây dựng thấp và không gian lưu trữ nhỏ cũng như khả năng tốt để nắm bắt cấu trúc cho dữ liệu đa tạp cục bộ và xử lý trên các tập dữ liệu lớn.

1.6 Xếp hạng đa tạp hiệu quả và vấn đề tra cứu ảnh

Để khắc phục hạn chế của xếp hạng đa tạp, trong [42] Bin Xu và các cộng sự đã đề xuất phương pháp xếp hạng đa tạp hiệu quả. Phương pháp này tập trung giải quyết hai vấn đề chính:

1- Xây dựng đồ thị neo (Anchor) có khả năng mở rộng thay cho đồ thị *k-NN* truyền thống.

2- Xây dựng mô hình tính toán xếp hạng hiệu quả, có chi phí tính toán nhỏ bằng cách thiết kế hình thức mới của ma trận kề.

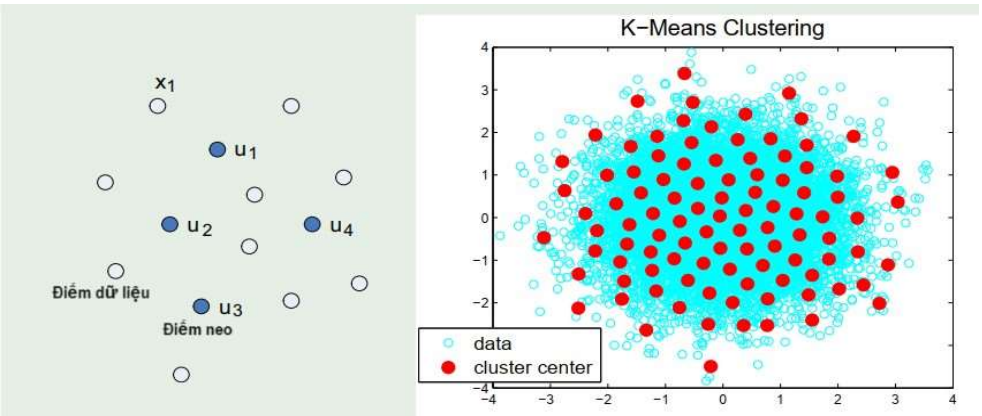
Mô hình này có hai pha tách rời: thứ nhất là pha ngoại tuyến (offline) để xây dựng đồ thị điểm neo cho toàn bộ cơ sở dữ liệu; pha thứ hai là pha trực tuyến (online) để xử lý một truy vấn mới.

Khái niệm điểm neo (Anchor): Là một điểm đại diện được chọn từ dữ liệu gốc để xây dựng một đồ thị neo (anchor graph) [42]. Các điểm neo sẽ đại diện cho các cụm dữ liệu tương tự hoặc có tính chất gần giống nhau và không chồng lấn lên nhau.

Do vậy điểm neo có một số đặc điểm sau:

+ Đại diện cho không gian dữ liệu: Các điểm neo được chọn sao cho chúng có khả năng phản ánh cấu trúc và đặc trưng của toàn bộ tập dữ liệu.

+ Bảo toàn thông tin cấu trúc: Mặc dù số lượng các điểm neo ít hơn nhiều so với tập dữ liệu ban đầu, nhưng chúng vẫn cần bảo toàn thông tin cấu trúc quan trọng của dữ liệu. Điều này có nghĩa là mối quan hệ giữa các điểm dữ liệu (như khoảng cách hoặc sự tương tự) được phản ánh một cách chính xác thông qua các điểm neo (các điểm dữ liệu có chung một điểm neo thì có mối quan hệ với nhau).



Hình 1.10. Các điểm dữ liệu và điểm neo [41]

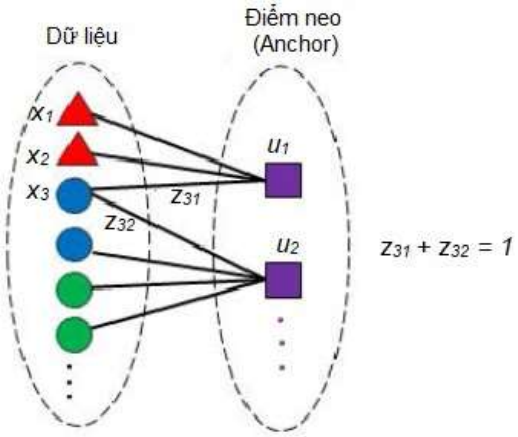
+ Dễ dàng cập nhật và mở rộng CSDL: Trong một số trường hợp, tập dữ liệu ban đầu có thể thay đổi theo thời gian, khi đó các điểm neo có thể được cập nhật hoặc điều chỉnh dễ dàng để phản ánh những thay đổi trong dữ liệu.

Do vậy điểm neo có thể phản ánh cấu trúc tổng thể của dữ liệu, được sử dụng như một điểm tham chiếu để ước lượng và hiểu biết về các mối quan hệ và thông tin trong tập dữ liệu. Hình 1.10 mô tả điểm neo và CSDL.

Xét tập dữ liệu $X = \{x_1, \ldots, x_n\} \in R^{n \times m}$ với n mẫu trong không gian m chiều và $U = \{u_1, \ldots, u_d\} \in R^{d \times m}$ (d<<n) là tập các điểm neo (Anchor) phân bố trong cùng không gian tương tự với tập dữ liệu. Ta định nghĩa $r: X \to R$là một hàm giá trị thực để gán mỗi điểm dữ liệu trong χ với một nhãn ngữ nghĩa, với mục đích là tìm ma trận trọng số $Z \in R^{n \times d}$ để biểu diễn mối quan hệ giữa các điểm dữ liệu trong $x_i \in X$và các điểm neo trong $u_k \in U$. Các giá trị *r(x)* được ước lượng cho mỗi điểm dữ liệu như là một trung bình trọng số của các nhãn trên Anchor bởi công thức:

$$r(x_i) = \sum_{k=1}^{d} z_{ik} r\,(u_k), i = \overline{1,n} \tag{1.21}$$

với các ràng buộc $\sum_{k=1}^{d} z_{ik} = 1$ và $z_{ik} > 0$, trong đó z_{ik} biểu diễn trọng số giữa điểm dữ liệu x_i và điểm neo u_k. Hình 1.14 biểu diễn điểm neo và các điểm dữ liệu.



Hình 1.14. Biểu diễn quan hệ điểm neo và dữ liệu

Xây dựng đồ thị neo (anchor graph) là cách tính vector trọng số z_i cho mỗi điểm dữ liệu x_i phụ thuộc vào hai vấn đề chính:

1) Các phương pháp xác định điểm neo và số lượng của neo;

2) Chi phí tính toán xây dựng đồ thị neo (tính vector trọng số).

Thông thường các điểm neo được xác định bằng các phương pháp như: học chủ động [37], lựa chọn ngẫu nhiên hoặc phương pháp phân cụm là các giải pháp đáng kể cần lựa chọn. Trong các nghiên cứu gần đây [36, 38, 39] điểm neo chính là các tâm cụm thu được thông qua thuật toán phân cụm K-means trên toàn bộ cơ sở dữ liệu đặc trưng.

Để đo trọng số cục bộ giữa điểm dữ liệu x_i và các điểm neo u gần x_i nhất, trong [40], S. T. Roweis đã đề xuất hàm mục tiêu với ý nghĩa hình học rõ ràng:

$$\min_{z_i} \varepsilon(z_i) = \frac{1}{2} ||x_i - \sum_{s=1}^{|NN(x_i)|} u_{s \in NN(x_i)} z_{is}||^2 \quad (1.22)$$

với điều kiện: $\sum_s z_{is} = 1, z_i \geq 0$

NN(x_i) là tập chỉ số của *s* điểm neo gần x_i nhất. Đây là dạng bài toán "*ước lượng trọng số cục bộ*". Bài toán này có thể được giải bằng cách sử dụng quy hoạch toàn phương chuẩn (quadratic programming-QP), tuy nhiên việc tính toán QP rất tốn nhiều thời gian. Do đó một số phương pháp đã được đề xuất để giải phương trình trên đó là phương pháp dựa trên phép chiếu Gradient [41] được đề xuất để tính toán ma trận trọng số và phương pháp hồi quy hạt nhân (Kernel Regression) [42] đã được áp dụng.

Trong [41] Liu và các công sự đã đề xuất các nguyên tắc để tính ma trận trọng số *Z* và ma trận kề *W* cho việc xây dựng đồ thị neo. Trong [41, 43] trọng số đo độ tương tự giữa x_i và u_j được tính bằng công thức:

$$Z_{ij} = \frac{\delta_i(j) K(x_i, u_j)}{\sum_{k=1}^{K} \delta_i(k) K(x_i, u_k)} \quad (1.23)$$

Trong công thức (1.23) thì $\delta_i \in \{0,1\}^K$ là tập vector chỉ số, với $\delta_i(j) = 1$ nếu u_j là điểm neo thứ *s* gần x_i nhất. Tập các $U = \{u_j\}_{j=1}^{K}$của *K* điểm neo được

khởi tạo bằng phép toán phân cụm *K-means* (tâm các cụm là các điểm neo).

Xây dựng ma trận kề W giữa các điểm dữ liệu được tính theo công thức:

$$W = ZD^{-1}Z^T \tag{1.24}$$

Với $D_{kk} = \sum_{i=1}^{n} Z_{ik}$ (là ma trận đường chéo, là tổng hàng k của Z)

Phần chứng minh công thức (1.23) được trình bày ở PL2.

Thuật toán 1.3: Thuật toán xếp hạng đa tạp hiệu quả (EMR)

Input:

- Tập dữ liệu $X = \{x_1, x_2 \ldots\ldots, x_n\} \subset \mathbb{R}^m$.

- Truy vấn q (có thể nằm ngoài tập dữ liệu)

- Số neo $d \ll$n (ví dụ: chọn bằng k-means)

- Kích thước lân cận s (số neo gần nhất cho mỗi điểm)

- Tham số $\propto$

Output: Véc tơ điểm xếp hạng $r^* \in \mathbb{R}^m$

Bước 1. Xây dựng đồ thị neo:

- Chọn tập neo $U = \{u_1, u_2 \ldots\ldots, u_d\}$ (ví dụ: chọn bằng k-means)

- Tính ma trận trọng số $Z \in R^{n \times d}$

$$z_{ki} = \frac{K\left(\frac{|x_i - u_k|}{\lambda}\right)}{\sum_{l=1}^{d} K\left(\frac{|x_i - u_k|}{\lambda}\right)}, \quad \lambda = \left|x_i - u_{[s]}\right| \tag{1.25}$$

Với Kernel Epanechnikov: $K(t) = \frac{3}{4}(1 - t^2)$ nếu $|t| \leq 1$, ngược lại 0.

Bước 2. Tính ma trận kề rút gọn;

$$W = Z^T Z$$

Tính D (đường chéo): $D_{ii} = Z_i^T v$, với $v = \sum_{j=1}^{n} z_j$

Bước 3. Xử lý truy vấn q:

Nếu q nằm ngoài tập dữ liệu, thêm cột z_q vào Z:

$$z_{kq} = \frac{K\left(\frac{|q - u_k|}{\lambda}\right)}{\sum_{l=1}^{d} K\left(\frac{|q - u_l|}{\lambda}\right)} \tag{1.27}$$

Bước 4. Tính điểm xếp hạng:

Đặt $H = ZD^{-1/2}$

Dùng công thức hiệu quả:

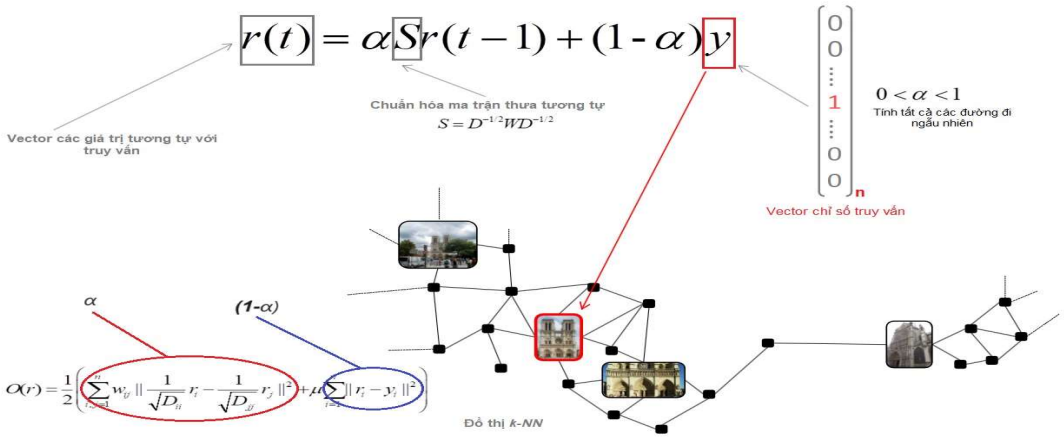
$$r^* = \left(I_n - H^T\left(HH^T - \frac{1}{\alpha}I_d\right)^{-1}H\right)y \quad (1.28)$$

Return

Nhận xét: Các phương pháp lựa chọn điểm neo và số lượng điểm neo được đề xuất trong các nghiên cứu [36, 38-39, 42, 44] chính là các tâm cụm thu được của thuật toán phân cụm K-means; Việc xây dựng đồ thị neo chính là bài toán ước lượng trọng số cục bộ giữa mỗi điểm dữ liệu và các điểm neo láng giềng có thể sai lệch và tốc độ tính toán chậm.

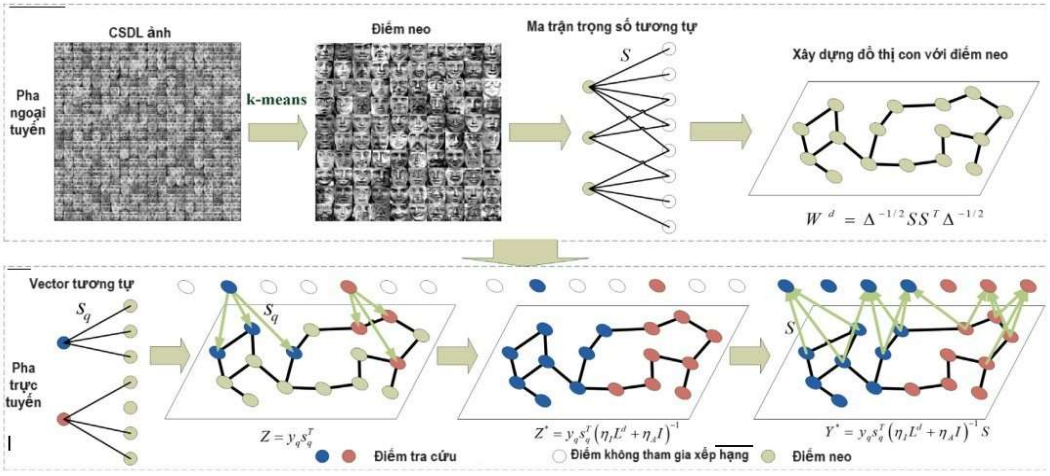
Các hệ thống CBIR thường sử dụng phương pháp so khớp từng cặp bằng cách sử dụng các độ đo khoảng cách như Euclid, Mahalanobis, Cosine,… để đo sự tương tự giữa ảnh truy vấn và mỗi ảnh cơ sở dữ liệu. Mặt khác, hệ thống CBIR dựa trên EMR là phương pháp khai thác các mối quan hệ liên quan (NN-Nearest Neighbor) giữa vector đặc trưng ảnh truy vấn tất cả các vector đặc trưng ảnh dữ liệu trong không gian đặc trưng nhất định, đồng thời xếp hạng điểm số của các ảnh được dán nhãn để ảnh được gán nhãn lan truyền đến ảnh không có nhãn qua một đồ thị có trọng số [45, 46]. Trong các nghiên cứu [36, 47-50] đã chứng minh được hiệu quả của EMR trong CBIR với biểu diễn ảnh bằng kết hợp các đặc trưng mức thấp. Mặt khác trong không gian các đặc trưng trực quan mức thấp của ảnh có thể là một đa tạp [44, 51] do vậy hình ảnh liên quan có xu hướng hình thành các cụm nào đó trong không gian đặc trưng; hình ảnh không thích hợp có thể hình thành một số cụm dữ liệu khác nhau với ngữ nghĩa khác nhau. Giá trị xếp hạng có thể được coi như một phép đo khoảng cách đa tạp, có ý nghĩa hơn để đối sánh mức độ liên quan ngữ nghĩa.

Tra cứu ảnh sử dụng thuật toán MR trong CBIR:



Hình 1.15. Quá trình tra cứu trong MR với đồ thị *K-NN*

Các thuật toán MR đã chứng minh được hiệu quả trong CBIR trên các tập CSDL nhỏ với các điểm dữ liệu truy vấn ở trong CSDL, khi phát sinh một truy vấn ở ngoài CSDL các thuật toán MR chưa xử lý được và cho hiệu quả tra cứu rất thấp. Để thực hiện được trên các tập CSDL lớn và có thể mở rộng CSDL và tra cứu được theo thời thực, trong các nghiên cứu [39, 45, 49, 79] đã đề xuất sử dụng xếp hạng đa tạp hiệu quả trong CBIR. Các mô hình đề xuất chia hệ thống thành hai pha như hình 1.16. Hệ thống CBIR sử dụng xếp hạng đa tạp hiệu quả đã chứng minh được hiệu quả tra cứu trên CSDL lớn và có thể thực hiện theo thời gian thực và độ chính xác cũng được cải thiện.



Hình 1.16. Mô hình hệ thống CBIR với SGR [39]

1.7 Phương pháp đánh giá hiệu quả trong CBIR

Trong tra cứu ảnh theo nội dung việc đánh giá hiệu năng là công việc quan

trọng. Trong [52] Muller-Henning và các cộng sự đưa ra độ đo lường hiệu năng của các hệ thống CBIR thường sử dụng như là độ chính xác (*precision (Pr)*); độ triệu hồi (*recall (Re)*). Nếu số các ảnh liên quan đến ảnh truy vấn cụ thể là A_Q, số các ảnh không liên quan đến truy vấn cụ thể là B_Q, tổng số các ảnh được truy vấn D_Q và số các ảnh liên quan được truy vấn E_Q.

Độ chính xác là tỉ số giữa các ảnh liên quan được truy vấn với tổng số các ảnh truy vấn, được tính theo công thức: $Pr = \frac{E_Q}{D_Q}$

Độ triệu hồi là tỉ số giữa số ảnh liên quan được truy vấn với toàn bộ số ảnh có liên quan trong CSDL: $Re = \frac{E_Q}{A_Q}$

Độ chính xác trung bình ARP (Average Retrieval Precision) thường được sử dụng để đánh giá độ chính xác của phương pháp được sử dụng trong CBIR. Hiệu quả truy vấn chung của một hệ thống được đo bằng trung bình tất cả độ chính xác. ARP được tính toán như sau:

$$ARP = average\left(\sum P_i\right) \tag{1.29}$$

Với P_i là độ chính xác của mỗi truy vấn. Nó là một độ đo hiệu quả để biểu diễn hiệu suất của hệ thống CBIR. Trong các thực nghiệm ở chương 2 và chương 3, luận án sử dụng độ chính xác trung bình để đánh giá hiệu quả của các phương pháp.

Với trường hợp nhiều truy vấn, độ chính xác trung bình trên tất cả các truy vấn ký hiệu là *MAP* (Mean Average Precision) được định nghĩa như sau:

$$MAP = \frac{1}{Q}\sum_{q=1}^{Q} AP_q \tag{1.30}$$

với Q là tổng số ảnh truy vấn.

Hầu hết các nghiên cứu trước đây chỉ đề cập đến giá trị trung bình của MAP, nhưng quá trình thử nghiệm cũng chỉ ra rằng mặc dù MAP có thể cao, hình ảnh có kết quả tra cứu kém vẫn tồn tại. Do đó, luận án đã thêm các công

thức: minP, σP [CT1, CT2] để xác định độ chính xác tối thiểu và tính đồng đều của kết quả tra cứu đối với tất cả các hình ảnh trong Q. Chỉ số minP rất quan trọng để kiểm tra và loại bỏ các trường hợp bị suy giảm gây ra kết quả kém. Giá trị minP được cải thiện đánh giá hiệu quả phục hồi trong các truy vấn cho kết quả xấu nhất.

$$minP = \min_{1 \le j \le |Q|} \left(\frac{m_j}{N}\right) * 100 \quad (1.31)$$

Giá trị σP nhỏ chứng tỏ rằng độ chính xác là đồng đều và tốt cho tất cả các hình ảnh truy vấn trong Q.

$$\sigma P = \sigma \left\{\frac{m_j}{N} * 100\right\}_{1 \le j \le Q} \quad (1.32)$$

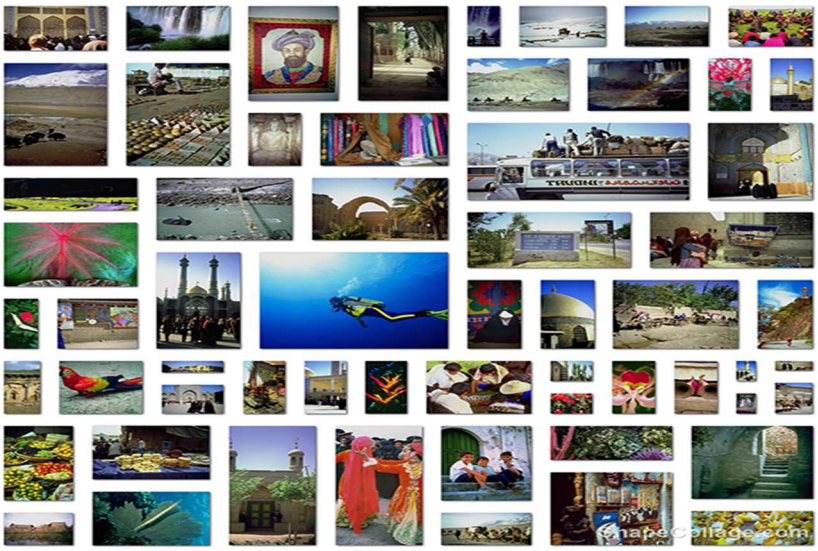
Tổng quan về các độ đo đánh giá hệ thống CBIR có thể tìm thấy trong [53]. Để thể hiện ý nghĩa thống kê, hiệu quả tra cứu được tính trên số lượng các truy vấn. Thông thường, số lượng các truy vấn là từ 100 đến 1000 [54]. Kích thước của tập dữ liệu ảnh là từ 1000 đến 20.000 [55]. Với các hệ thống tra cứu quy mô lớn, kích thước tập dữ liệu ảnh có thể lên đến 80 triệu [56]

1.8 Một số CSDL thực nghiệm cho tra cứu ảnh

Luận án tiến hành các thực nghiệm trên các tập dữ liệu (1) Sign300HD, (2) Corel30K, (3) Leaf30 và (4) VGG-60K [57].

Tập dữ liệu thứ nhất: Bộ sưu tập hình ảnh gốc Corel Photo Gallery [58] nổi tiếng trong cộng đồng nghiên cứu bao gồm hơn 800 đĩa CD, mỗi đĩa chứa các ảnh của một khái niệm chủ đề cụ thể tiền cảnh nổi bật. Các khái niệm chủ đề này có thể rất rộng như "đại dương, mùa thu" hoặc có thể giới hạn phạm vi nhỏ hơn như "hoàng hôn, đường cao tốc". Do sự phức tạp của các chủ đề ảnh thay đổi từ loại này sang loại khác và cỡ của bộ sưu tập là rất lớn nên toàn bộ tập ảnh thường không được sử dụng trong một hệ thống tra cứu cụ thể. Điều đó dẫn đến tình trạng mỗi nhóm nghiên cứu tạo ra bộ Corel con của riêng mình trong các thực nghiệm của họ.

Để đánh giá độ chính xác của các phương pháp tra cứu, nhiều hệ thống CBIR sử dụng một tập con của Corel Photo Gallery gồm 31.695 ảnh làm tập dữ liệu ảnh thử nghiệm [58] gọi là tập Corel30K. Tập Corel30K bao gồm 306 chủ đề khác nhau và mỗi chủ đề thường có 100 ảnh trừ một số ít chủ đề có nhiều hơn 100 ảnh. Kích thước của mỗi ảnh trong mỗi chủ đề cố định là 256×384 hoặc 384×256. Một số ảnh minh họa trình bày như Hình 1.17



Hình 1.17: Một số hình ảnh trong tập CSDL Corel30K

Tập dữ liệu thứ hai Leaf30 [59], gồm 11600 hình ảnh lá cây của 30 loài cây khác nhau và được đặt trong 58 nhóm, với mỗi nhóm gồm 200 ảnh của mỗi loại lá, trong đó có chứa hình ảnh lá khỏe mạnh, lá sâu bệnh (từ 1 -5 loại bệnh). Hình 1.18 chỉ ra một số mẫu ảnh trong tập dữ liệu này.

Hình 1.18: Một số hình ảnh trong tập CSDL Leaf30

Tập dữ liệu thứ ba Sign300HD [59], gồm 4370 hình ảnh chữ ký của 300 người. Bộ dữ liệu này được tổng hợp từ nhiều nguồn, bao gồm: GPDS 1-150, ICEDAR, Bengali và thu thập từ chữ ký thực tế của 50 người tại Trường Đại học Hồng Đức. Đây là bộ dữ liệu phức tạp và chưa gắn nhãn. Tập dữ liệu có cấu trúc 300 lớp, với 16 ảnh trên một lớp, có dung lượng 285 MB. Ảnh chữ ký được đánh số từ 01-16 trong mỗi thư mục. Hình 1.19 chỉ ra một số mẫu ảnh trong tập dữ liệu này.

Hình 1.19: Một số hình ảnh trong tập CSDL SIGN300HD

Cuối cùng tập dữ liệu VGG-60K [57] là tập con của tập dữ liệu hình ảnh để nhận dạng khuôn mặt VGGFACE2 với 169396 ảnh trong 500 lớp. Tập VGG-60K bao gồm 60.000 hình ảnh của 500 người được lấy ngẫu nhiên từ bộ dữ liệu VGGFace2. Bộ dữ liệu được chia thành 500 lớp, mỗi lớp chứa 120 ảnh, có dung lượng 768 MB. Tên hình ảnh được gán bởi tên thư mục với số thứ tự hình ảnh trong thư mục... Hình 1.20 chỉ ra một số mẫu ảnh trong tập này. Ảnh trong các tập dữ liệu có kích thước, độ phân giải và mầu sắc khác nhau.

Hình 1.20: Một số hình ảnh trong tập CSDL VGG-60K

Các tập dữ liệu này được tổ chức thành các lớp ngữ nghĩa theo cách con người nhận thức về độ tương tự. Mỗi lớp biểu diễn một chủ đề ngữ nghĩa khác nhau, các ảnh trong cùng một lớp được xem là liên quan. Bảng 1.3 mô tả số lượng và số lớp ảnh trong các tập dữ liệu thực nghiệm.

Bảng 1.5. Các tập dữ liệu ảnh

Tập ảnh	Số lượng ảnh	Số lớp ảnh
Leaf30	11600	116
Sign300HD	4370	300
Corel30K	31695	306
VGG-60K	60000	500
Dermnet	5888	23

Về mặt thuật toán, các hệ thống CBIR chủ yếu tập trung xử lý từng ảnh dựa trên đặc trưng hình ảnh (đặc trưng mức thấp, mức cao) và thông tin metadata, mà không phụ thuộc vào dữ liệu mô tả của người dùng. Do đó, có thể xem xét các ảnh trong tập dữ liệu ảnh mà không cần khai thác thêm thông tin bổ sung từ bên ngoài.

Tuy nhiên, trong các hệ thống CBIR triển khai thực tế như Google Lens, Bing Image Search… ảnh thường đi kèm với thông tin chuyên đề và được lưu trữ trong một cơ sở dữ liệu ảnh (CSDL ảnh). Để đảm bảo tính nhất quán và thuận tiện trong trình bày, luận án này sử dụng đồng thời hai thuật ngữ tập dữ liệu ảnh (dataset ảnh) và CSDL ảnh, mà không quá nhấn mạnh sự khác biệt giữa chúng.

1.9 Kết luận Chương 1

Trong chương 1 luận án đã trình bày về tra cứu ảnh dựa trên nội dung, các độ đo khoảng cách, độ đo tương tự, tiếp cận học độ đo khoảng cách, độ đo tương tự là thành phần quan trọng trong các hệ thống CBIR.

Cũng ở trong chương này luận án tập trung giới thiệu và phân tích các xếp hạng EMR cho đặc trưng mức thấp, xếp hạng EMR cho đặc trưng mức cao (bao gồm: đặc trưng CNN, đặc trưng véc tơ nhúng), qua đó rút ra một số kết luận về sử dụng EMR cho CBIR như sau:

- Ưu điểm

+ Vừa có tính địa phương: lấy theo lân cận gần nhất; vừa có tính toàn cục: có tính lan tỏa giữa các lân cận.

+ Đưa ra được độ tương tự giữa ảnh truy vấn và ảnh trong CSDL có hiệu quả khá tốt.

+ Nhược điểm

+ Dù đã cải tiến từ MR sang EMR tuy nhiên với hệ thống ảnh lớn thì số anchor lớn nhưng hệ thống vẫn chậm. Với phép toán ma trận tính toán số lượng anchor lớn thì rất chậm.

+ Nếu thêm ảnh mới vào cơ sở dữ liệu thì có khả năng phải xây dựng lại mô hình EMR (tính lại ma trận Z).

+ Chỉ xác định được ảnh truy vấn và ảnh trong cơ sở dữ liệu như vậy sẽ không tính được độ tương tự ảnh của 2 ảnh nằm ngoài cơ sở dữ liệu.

Từ đó, Luận án đưa ra 02 vấn đề cần giải quyết.

Vấn đề 1: Xây dựng bộ xếp hạng CSDL ảnh theo ảnh truy vấn dựa trên nhiều bộ xếp hạng EMR.

(Có rất nhiều loại đặc trưng ảnh (biểu diễn ảnh), nếu chỉ dùng một xếp hạng EMR thì không khai thác được điểm mạnh của từng loại đặc trưng, vì vậy cần kết hợp nhiều bộ xếp hạng EMR. Ở đây ta xem ảnh như một biểu diễn bởi nhiều bộ đặc trưng, xây dựng độ đo khoảng cách, độ đo tương tự của các cặp ảnh dựa trên các độ đo khoảng cách, độ đo tương tự của từng bộ đặc trưng.)

Vấn đề 2: Xây dựng một độ đo tương tự ảnh được học dựa trên kết quả xếp hạng của EMR.

Xây dựng độ đo tương tự trên đặc trưng ảnh (chỉnh trọng số của các thành phần đặc trưng) sau đó sử dụng dữ liệu ảnh đầu vào và học để tạo ra 1 véc tơ tương ứng với ảnh đầu vào (véc tơ nhúng).

Chương 2 và Chương 3 của luận án này sẽ tập trung giải quyết các vấn đề 1 và vấn đề 2 để nâng cao độ chính xác cho các hệ thống CBIR.

CHƯƠNG 2: PHƯƠNG PHÁP TRA CỨU ẢNH SỬ DỤNG THUẬT TOÁN KẾT HỢP NHIỀU BỘ XẾP HẠNG ĐA TẠP HIỆU QUẢ

Trong Chương 1 đã giới thiệu và phân tích một số kỹ thuật được sử dụng trong CBIR và xếp hạng đa tạp trong CBIR. Để nâng cao độ chính xác trong CBIR, từ các phân tích trong Chương 1, luận án nhận thấy EMR hiệu quả trong việc khai thác cấu trúc phi tuyến của dữ liệu ảnh, vì vậy trong Chương 2 luận án đề xuất phương pháp kết hợp nhiều bộ xếp hạng EMR (CoEMR), đánh giá hiệu quả của thuật toán CoEMR đề xuất trên một số tập dữ liệu có dạng "đa tạp" với các chỉ số khác nhau. Chương 2 cũng đưa ra thực nghiệm cho hệ thống CBIR trên các tập dữ liệu VGG60K, LEAF30… Nội dung những đề xuất của chương này là kết quả đã được công bố tại công trình nghiên cứu [CT1-CT4].

Các phương pháp xếp hạng trên đồ thị như: AGR[38], SGR [39], MR [33-35] và EMR [36, 38,39,42] đã được áp dụng rộng rãi trong việc tra cứu thông tin và được chứng minh là có hiệu suất cao trên nhiều loại dữ liệu khác nhau. Đặc biệt, trong các nghiên cứu [59, 116, 127] MR, EMR đã được áp dụng trong CBIR. Tuy nhiên, MR có hạn chế khi xử lý các cơ sở dữ liệu quy mô lớn - đòi hỏi chi phí tính toán cao, trong cả giai đoạn xây dựng đồ thị và tính toán xếp hạng. Để giải quyết những hạn chế của MR, EMR các nghiên cứu như: tra cứu ảnh dựa trên mở rộng xếp hạng đa tạp SMR (Scaling manifold ranking based image retrieval) [33], Tra cứu ảnh dựa vào xếp hạng đa tạp tổng quát - GMR-Generalized manifold-ranking-based image retrieval [39], Xếp hạng đa tạp nhanh cho tra cứu ảnh dựa vào nội dung - FMR(Fast manifold-ranking for content-based image retrieval) [41] … các phương pháp trên đều dựa vào xếp hạng đa tạp tổng quát véc tơ đặc trưng mức thấp hoặc mức với các điểm dữ liệu, điểm neo được xây dựng dựa vào đồ thị *K-NN* và ổn định toàn cục. Tuy nhiên các phương pháp này chỉ áp dụng cho đặc trưng mức thấp hoặc mức cao và chưa xử lý các tình huống tra cứu với điểm truy vấn bên ngoài CSDL.

Do vậy, các phương pháp cải tiến đã kết hợp đặc trưng mức thấp và mức cao như: Tra cứu hình ảnh dựa trên nội dung bằng các kết hợp thống tin đặc trưng của hình ảnh - Content based image retrieval using image features information fusion [118], Sử dụng xếp hạng đa tạp với đa đặc trưng trong tra cứu ảnh - MMR (Multi-Manifold Ranking: Using Multiple Features for Better Image Retrieval) [113], Tra cứu hình ảnh hiệu quả dựa trên lựa chọn điểm neo trong EMR với sự kết hợp đặc trưng thấp và mức cao - (Efficient content-based image retrieval based on anchor point selection in manifold ranking with combined CNN and low-level features) [128], Tăng cường phản hồi liên quan dài hạn trong CBIR với tối ưu hóa mở rộng độ thị con - SRG (A scalable sub-graph regularization for efficient content based image retrieval with long-term relevance feedback enhancement) [122], các phương pháp đề xuất nhằm khắc phục hạn chế của việc sử dụng riêng rẽ từng đặc trưng và thực hiện tính toán xếp hạng được trên các tập CSDL lớn, các phương pháp này chủ yếu tập trung hướng sử dụng nhiều đặc trưng ảnh trong xếp hạng EMR nhằm tăng cường hiệu qua tra cứu ảnh. Mặc dù các tiếp cận này là khá hiệu quả nhưng do chỉ sử dụng một bộ xếp hạng nên các phương pháp này vẫn tồn tại hạn chế là không tận dụng được các điểm mạnh của từng bộ đặc trưng ảnh (không linh hoạt trong việc điều chỉnh tham số để tăng độ chính xác xếp hạng), ngữ nghĩa của ảnh và tăng độ phức tạp tính toán (chiều của véc tơ đặc trưng rất cao).

Trong chương này luận án đề xuất kỹ thuật tổ hợp xếp hạng EMR thứ nhất cho đặc trưng mức thấp và EMR thứ hai cho đặc trưng mức cao và xây dựng mô hình tra cứu ảnh dựa vào thuật toán CoEMR.

2.1 Tiếp cận kết hợp đặc trưng mức thấp và đặc trưng CNN trong mô hình CoEMR đề xuất

Trong phần này, luận án sẽ giới thiệu một phương pháp kết hợp mới gọi là CoEMR nhằm cải thiện hiệu quả tra cứu ảnh thông qua việc sử dụng đồng

thời các đặc trưng mức thấp và đặc trưng CNN. CoEMR dựa trên việc áp dụng hai xếp hạng EMR riêng biệt: một xếp hạng cho các đặc trưng mức thấp và một xếp hạng khác cho các đặc trưng CNN. Sau khi các ảnh trong cơ sở dữ liệu được đánh giá và xếp hạng theo từng đặc trưng, kết quả từ hai xếp hạng này sẽ được kết hợp lại để tạo ra một thứ hạng mới cho từng ảnh trong cơ sở dữ liệu so với ảnh truy vấn. Cách tiếp cận này không chỉ tận dụng tối đa thông tin từ cả hai bộ đặc trưng mà còn giúp nâng cao độ chính xác của hệ thống truy vấn hình ảnh, đảm bảo rằng các kết quả trả về phản ánh đúng mức độ tương tự giữa các ảnh một cách toàn diện hơn. Phần trình bày này sẽ đi sâu vào chi tiết từng bước của phương pháp kết hợp CoEMR, từ quá trình xếp hạng ban đầu đến cách thức kết hợp kết quả để tạo ra xếp hạng cuối cùng cho các ảnh trong cơ sở dữ liệu.

2.2 Phát biểu các ràng buộc cho lớp hàm kết hợp xếp hạng

Đối với một ảnh tra cứu I_q và CSDL E chứa n ảnh, mỗi ảnh I trong E sẽ có 2 thứ hạng a và b dựa trên các mô hình xếp hạng EMR khác nhau (chẳng hạn như EMR_{LF} cho đặc trưng mức thấp và EMR_{HF} cho đặc trưng mức cao), khi kết hợp lại sẽ tạo ra một thứ hạng mới của ảnh I so với I_q.

Để kết hợp xếp hạng 2 véc tơ xếp hạng CSDL ảnh theo ảnh truy vấn, luận án sử dụng một lớp hàm CB kết nhập 2 số thực (có thể âm) thành một số thực duy nhất. Chúng ta có thể phát biểu dạng tổng quát cho hàm kết hợp 2 xếp hạng $r = CB(a, b)$ ở đây CB là hàm 2 biến số thỏa mãn:

Yêu cầu 1. $\min(a, b) \leq CB(a, b) \leq \max(a, b)$, với mọi số a,b

Yêu cầu 2. CB là không giảm theo nghĩa:

$CB(a_1, b_1) \geq CB(a_2, b_2)$, với mọi $a_1 \geq a_2$, $b_1 \geq b_2$.

Các hàm CB được sử dụng trong luận án bao gồm:

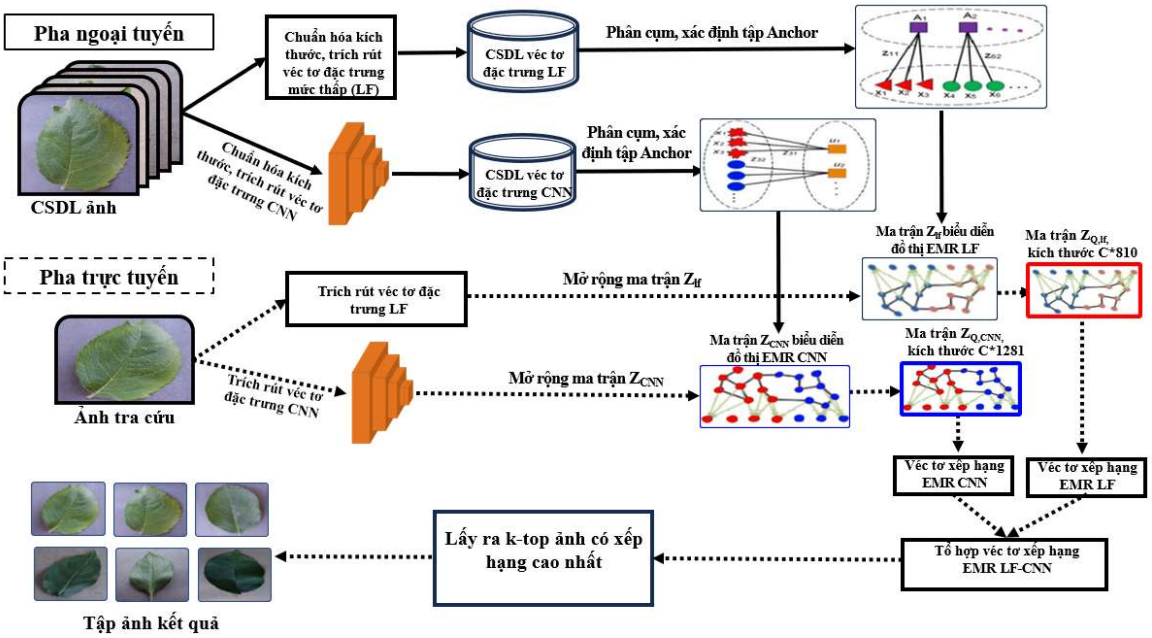
- Kết hợp tuyến tính $CB(r_1, r_2) = \{\alpha r_1 + \beta r_2\}$. $(\alpha, \beta > 0, \alpha + \beta = 1)$ (2.1)

- Kết hợp kiểu lựa chọn: $CB(r_{1,}r_2) = \begin{cases} r_1 \\ r_2 \end{cases} \quad if\ r_2 \geq th$ (2.2)

- Kết hợp dạng luỹ thừa bậc lẻ:

$$CB(r_{1,}r_2) = \begin{cases} \sqrt[3]{\frac{\{r_1^3 + r_2^3\}}{2}} \\ \min(r_1, r_2) \end{cases} \quad if\ r_1 + r_2 \geq 0 \quad (2.3)$$

Mô hình đề xuất hệ thống CBIR sử dụng phương pháp kết hợp các bộ xếp hạng EMR được mô tả tại hình 2.1.



Hình 2.1. Mô hình hệ thống đề xuất CBIR với việc sử dụng kết hợp xếp hạng Co*EMR*

Kiến trúc đề xuất hệ thống CBIR trong hình 2.1 bao gồm các bước sau:

+ Pha ngoại tuyến:

1. Cơ sở dữ liệu ảnh được trích rút các đặc trưng mức thấp *(low-level feature),* đồng thời sử dụng mô hình tiền huấn luyện để trích xuất véc tơ đặc trưng mức cao (đặc trưng véc tơ nhúng [CT1], đặc trưng CNN [122]);

2. Chuẩn hóa các vector đặc trưng mức thấp [60].

3. Xây dựng xếp hạng EMR cho bộ đặc trưng mức thấp, xây dựng xếp

hạng EMR cho đặc trưng mức cao (véc tơ nhúng, CNN).

4. Kết hợp xếp hạng EMR cho đặc trưng mức thấp và xếp hạng EMR cho đặc trưng mức cao theo ba (03) phương pháp: Kết hợp tuyến tính, kết hợp kiểu lựa chọn, kết hợp dạng lũy thừa bậc lẻ.

+ Pha trực tuyến:

Ảnh truy vấn được trích rút đặc trưng mức thấp, véc tơ nhúng và sử dụng xếp hạng EMR đối sánh các ảnh tương tự và đưa ra kết quả truy vấn [CT1, CT2, CT4], cụ thể:

Cấu trúc đồ thị neo trong EMR được xây dựng với các bước sau:

Bước 1: Xây dựng C điểm neo: Chọn C điểm từ không gian dữ liệu hoặc có thể thông qua một quá trình lựa chọn ngẫu nhiên hoặc một phương pháp tối ưu nào đó. Điểm neo này sẽ đóng vai trò là các điểm tham chiếu trong không gian dữ liệu.

Bước 2: Tìm s điểm neo gần nhất với mỗi điểm dữ liệu x_i: Đối với mỗi điểm dữ liệu trong tập X, xác định s điểm neo gần nhất với nó, thường dựa trên một tiêu chí khoảng cách như khoảng cách Euclide.

Bước 3: Tìm trọng số hồi quy z_i của s điểm neo đối với mỗi điểm dữ liệu: Đối với mỗi điểm dữ liệu, tính các trọng số hồi quy z_i mà thông qua đó có thể biểu diễn mối quan hệ giữa các điểm dữ liệu dựa vào các điểm neo.

Bước 4: Xây dựng ma trận hồi quy Z: Ma trận Z được tạo ra từ các trọng số hồi quy của tất cả các điểm dữ liệu, thể hiện mức độ liên kết của từng điểm dữ liệu với các điểm neo của nó.

Bước 5: Xây dựng ma trận kề có trọng số W: Dựa trên ma trận hồi quy Z, xây dựng ma trận kề W để biểu diễn mối quan hệ và sự tương tác giữa các điểm dữ liệu ban đầu dựa vào các điểm neo.

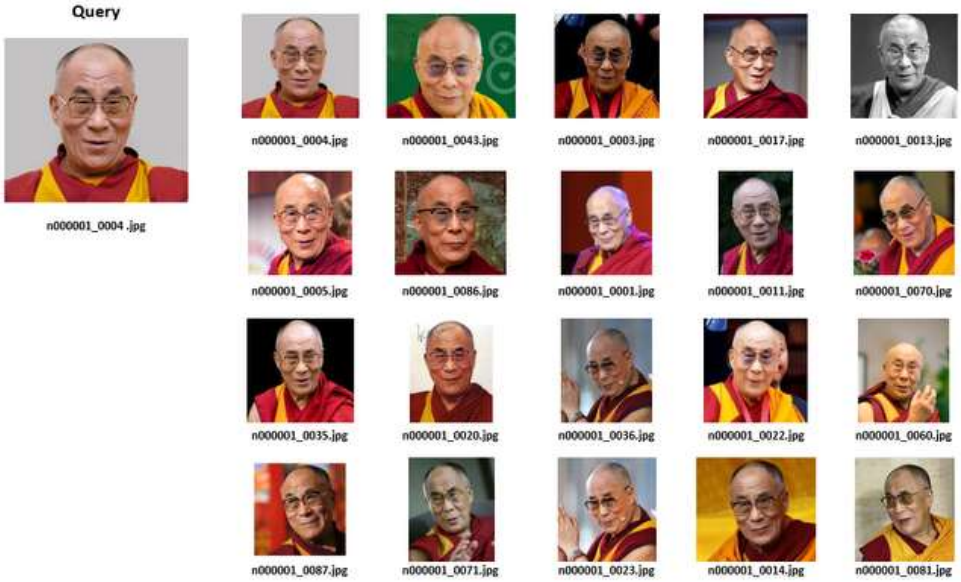
2.3 Kết hợp lựa chọn 2 xếp hạng EMR.

Ý tưởng đề xuất kết hợp lựa chọn xuất phát từ việc nhận thấy rằng có một

số loại đặc trưng CNN trên một lớp ảnh có độ phân biệt rất cao, ví dụ như các vector nhúng được xác định bởi mô hình FaceNet cho các ảnh khuôn mặt. Wang và cộng sự [38] đã phân tích và thử nghiệm về bộ dữ liệu ImageNet và chỉ ra rằng hai biểu diễn đặc trưng fc4096a và fc4096b, được trích xuất từ lớp thứ nhất và lớp thứ hai của AlexNet, có tính khái quát tốt hơn khả năng hơn các đặc trưng khác và mang lại hiệu suất cao.Trong các trường hợp này, đặc trưng CNN thường thể hiện hiệu quả vượt trội, mang lại độ chính xác cao trong tra cứu ảnh. Tuy nhiên, do cấu trúc phân tầng của mạng CNN, nên cũng có những trường hợp đặc thù mà việc sử dụng đơn thuần đặc trưng CNN lại không mang lại kết quả mong muốn [118].

Khi kết hợp với các đặc trưng mức thấp, phương pháp này có thể cải thiện đáng kể kết quả truy vấn cho những trường hợp mà đặc trưng CNN không đạt hiệu quả cao. Đặc trưng mức thấp như màu sắc, hình dạng, và kết cấu giúp bổ sung thông tin cần thiết, làm tăng khả năng phân biệt và nâng cao độ chính xác của hệ thống truy vấn. Ví dụ, nếu một ảnh truy vấn có kết quả rất kém khi chỉ sử dụng đặc trưng CNN, việc kết hợp với đặc trưng mức thấp có thể giúp cải thiện đáng kể kết quả truy vấn.

Hình 2.2 cho thấy rõ ràng việc những hình ảnh có độ chính xác cao khi sử dụng đặc trưng CNN vẫn gặp khó khăn trong một số trường hợp đặc biệt, nhưng khi kết hợp với đặc trưng mức thấp, kết quả truy vấn được cải thiện đáng kể. Điều này chứng tỏ rằng việc kết hợp lựa chọn không chỉ giúp tận dụng tối đa ưu điểm của cả hai loại đặc trưng mà còn mang lại một hệ thống truy vấn hình ảnh toàn diện và hiệu quả hơn. Phương pháp này sẽ được phân tích chi tiết trong các phần tiếp theo của luận án, bao gồm các bước thực hiện cụ thể và các kết quả thực nghiệm minh họa cho hiệu quả của phương pháp.



Hình 2.2. Kết quả tra cứu ảnh sử dụng đặc trưng CNN, cho kết quả tốt.

Hình 2.3. Kết quả tra cứu ảnh sử dụng đặc trưng CNN, ảnh có "cảm quan" tốt nhưng kết quả tra cứu lại kém.

Trong phần này luận án đề xuất kỹ thuật kết hợp kiểu lựa chọn 2 bộ xếp hạng EMR cho đặc trưng mức thấp, đặc trưng véc tơ nhúng và xây dựng mô hình tra cứu ảnh dựa vào thuật toán CoEMR-S.

Kết hợp CoEMR kiểu lựa chọn (CoEMR-S)

Giả sử rằng tập dữ liệu ảnh gốc, mỗi ảnh có một vector đặc trưng duy nhất tương ứng và với một ảnh truy vấn I_Q, hệ thống chọn xếp hạng theo hình ảnh đặc trưng mức thấp $r^*_{lf.v.Q,i}$ kết hợp với xếp hạng của hình ảnh đặc trưng vector nhúng $r^*_{EmbV.Q,i}$, trong đó các vector xếp hạng được tính bằng EMR.

Sau đây là thuật toán chi tiết mà luận án đề xuất.

Thuật toán 2.1. CoEMR-S (Thuật toán kết hợp lựa chọn 2 bộ xếp hạng EMR).

Input: $\{I_i\}_{1\le i\le n}$ là tập ảnh gốc, I_Q ảnh truy vấn.

M*N là kích thước ảnh chuẩn hóa cho huấn luyện theo thuật toán FaceNet.

d: số chiều của một vector đặc trưng ảnh.

C: số lượng điểm neo của thuật toán EMR, tham số a $\in$ (0,1) (a $\approx$ 1),

th: ngưỡng, th > 0.

Output: r^* là thứ hạng tương tự với ảnh I_Q của ảnh I_i trong CSDL ảnh E.

Bước 1 (ngoại tuyến): **Xây dựng đồ thị EMR [146]**

Bước 1a. Huấn luyện tập ảnh E theo mô hình FaceNet với kích thước chuẩn hóa ảnh đầu vào là M*N, thu nhận được:

1.1. Tham số model W_{EmbV}.

1.2. Chạy từng ảnh $\{I_i\}_{1\le i\le n}$ qua mô hình thu được tập các vector đặc trưng $\{v.E_i\}_{1\le i\le n}$ với chiều d_{EmbV}=128.

1.3. Xác định C điểm neo $\{A_c\}_{1\le c\le C}$ của tập ảnh $\{I_i\}_{1\le i\le n}$ dựa trên thuật toán Kmeans, sử dụng các đặc trưng vector nhúng $\{v.E_i\}_{1\le i\le n}$

1.4. Xác định ma trận kề W=(w_{ij}) của thuật toán EMR sử dụng các vector nhúng trích xuất từ mạng FaceNet $\{v.E_i\}_{1\le i\le n}$.

1.5. Xác định ma trận trọng số Z kích thước C*n của EMR.

Bước 1b. Trích rút đặc trưng mức thấp: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST, thu được tập véc tơ đặc trưng với số chiều

d_{LF}=809. Lặp lại bước 1.3 - 1.5 như ở bước 1a với đặc trưng mức thấp.

Bước 2 (trực tuyến): Kết hợp xếp hạng 2 EMR

2.1. Chuẩn hóa ảnh I_Q về kích thước M*N, chạy qua mô hình FaceNet đã xác định ở bước 1.1, thu nhận được vector đặc trưng vector nhúng v.E_Q với d chiều.

2.2. Mở rộng ma trận Z theo EMR [42] (có thể xem công thức (4) ở trên): Sử dụng C giá trị khoảng cách của E_Q với các phần tử neo của A_c, chúng ta thu được ma trận trọng số Z_Q mới có kích thước C*(n+1).

2.3. Đặt $r_Q = \{r_i\}_{1 \le i \le n+1}, r_i = 0 \forall i = \overline{1,n}, r_{n+1} = 1.0$. Từ ma trận trọng số Z_Q chúng ta xác định vector thứ hạng $r_{emb} = r^*_{EmbV.Q,i}$bằng thuật toán EMR [42].

Bước 3: Lặp lại bước 2.1 - 2.3 nhưng với đặc trưng mức thấp lf.E_i ta được giá trị xếp hạng là $r_{lf} = r^*_{lf.v.Q,i}$

Bước 4: **Kết hợp vector xếp hạng của 2 EMR:**

Kết hợp kiểu lựa chọn:

$$r^* = \{CB(r_{emb,i}, r_{lf,i})\}_{1 \le i \le n}, \text{ở đó } CB(a,b) = \begin{cases} a & if\ a \ge th \\ b & \end{cases}$$

Kết quả trả về: r^*.

Đánh giá độ phức tạp của thuật toán CoEMR-S

Thuật toán **CoEMR-S** chia thành 2 pha: offline, online và được thực hiện bằng cách tổ hợp 2 xếp hạng EMR.

Đối với pha offline, độ phức tạp là:

Phân cụm K-means, xác định $n \times C$ khoảng cách giữa từng tâm cụm và các điểm dữ liệu: *O(C*n*d)*.

Tính toán xếp hạng truy vấn: O(n*C+C^3)

trong đó C là số cụm, n số mẫu đầu vào, d số chiều véc tơ đặc trưng.

Đối với pha online, độ phức tạp là: *O(C*n*d)*

Vậy thuật toán **CoEMR-S** có độ phức tạp tương đương với độ phức tạp

của EMR gốc [42] là $O(dn + d^3)$.

Nhận xét: Thuật toán CoEMR-S cung cấp một cách kết hợp kiểu lựa chọn xếp hạng của EMR cho đặc trưng mức thấp và EMR cho đặc trưng vector nhúng một cách hiệu quả, bổ trợ kết quả cho nhau từ đó làm tăng độ chính xác của kết quả truy vấn hình ảnh trong CBIR.

2.4 Kết hợp tuyến tính 2 xếp hạng EMR.

Ý tưởng đề xuất kết hợp tuyến tính xuất phát từ việc nhận thấy rằng đối với ảnh chữ ký thì đặc trưng mô tả ảnh chữ ký là hạn chế, khi áp dụng bài toán nhận dạng ảnh chữ ký: đặc trưng mức thấp (HOG, LBP) mô tả tốt các đặc điểm của ảnh chữ ký như cạnh, góc, điểm nổi bật, trong khi đặc trưng CNN (SigNet) có khả năng học đặc điểm biên, hoặc mẫu hình học [133]. Luận án tìm cách đưa ra một mô hình kết hợp tuyến tính để có thể tận dụng được tính chất của toàn bộ các đặc trưng mức thấp, mức cao.

Điều này có thể đồng nghĩa với việc mô hình có khả năng học và tận dụng tối đa đặc trưng khác nhau từ dữ liệu ảnh chữ ký, mô tả cả tính cục bộ và tính toàn cục của ảnh, giúp cải thiện khả năng tổng quát hóa và độ chính xác của mô hình. Phương pháp kết hợp tuyến tính các bộ xếp hạng EMR bao gồm các bước thực hiện cụ thể và các kết quả thực nghiệm minh họa cho hiệu quả của phương pháp.

Trong phần này luận án đề xuất kỹ thuật kết hợp tuyến tính 2 bộ xếp hạng EMR cho đặc trưng mức thấp, đặc trưng CNN (SigNet) và xây dựng mô hình tra cứu ảnh dựa vào thuật toán CoEMR-L.

Kết hợp CoEMR kiểu tuyến tính (CoEMR-L)

Giả sử rằng trong tập dữ liệu ảnh gốc, mỗi ảnh có một vector đặc trưng duy nhất tương ứng và với một ảnh truy vấn I_Q, hệ thống chọn xếp hạng theo hình ảnh đặc trưng mức thấp $r^*_{lf.v.Q,i}$ kết hợp với xếp hạng của hình ảnh đặc trưng CNN $r^*_{CNN.Q,i}$, trong đó các vector xếp hạng được tính bằng EMR.

Sau đây là thuật toán chi tiết mà luận án đề xuất.

Thuật toán 2.2. CoEMR-L (Thuật toán kết hợp tuyến tính 2 bộ xếp hạng EMR).

Input: $\{I_i\}_{1\le i\le n}$ là tập ảnh gốc, I_Q ảnh truy vấn.

M*N là kích thước ảnh chuẩn hóa cho huấn luyện theo thuật toán EfficientNet.

d: số chiều của một vector đặc trưng ảnh.

C: số lượng điểm neo của thuật toán EMR, tham số a ∈ (0,1) (a ≈ 1),

α, β:trọng số kết hợp tuyến tính xếp hạng của 2 EMR, α, β >0, α + β = 1.

Output: r^* là thứ hạng tương tự với ảnh I_Q của ảnh I_i trong CSDL ảnh E.

Bước 1 (ngoại tuyến): **Xây dựng đồ thị EMR**

Bước 1a. Huấn luyện tập ảnh E theo mô hình SigNet với kích thước chuẩn hóa ảnh đầu vào là M*N, thu nhận được:

1.1. Tham số mô hình W_{CNN}.

1.2. Chạy từng ảnh $\{I_i\}_{1\le i\le n}$ qua mô hình thu được tập các vector đặc trưng $\{v.E_i\}_{1\le i\le n}$ với chiều d_{CNN}=256.

1.3. Xác định C điểm neo $\{A_c\}_{1\le c\le C}$ của tập ảnh $\{I_i\}_{1\le i\le n}$ theo thuật toán K-means, sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$

1.4. Xác định ma trận kề W=(w_{ij}) của thuật toán EMR sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$.

1.5. Xác định ma trận trọng số Z kích thước C*n của EMR.

Bước 1b. Trích rút đặc trưng mức thấp: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST, thu được tập véc tơ đặc trưng với số chiều d_{LF}=809. Lặp lại bước 1.3 - 1.5 như ở bước 1a với đặc trưng mức thấp.

Bước 2 (trực tuyến): Kết hợp xếp hạng 2 EMR

2.1. Chuẩn hóa ảnh I_Q về kích thước M*N, chạy qua mô hình SigNet đã xác định ở bước 1.1, thu nhận được vector đặc trưng CNN v.E_Q với d_{CNN} chiều.

2.2. Mở rộng ma trận Z theo EMR [42]: Sử dụng C giá trị khoảng cách của E_Q với các phần tử neo của A_c, chúng ta thu được ma trận trọng số Z_Q mới có kích thước C*(n+1).

2.3. Đặt $r_Q = \{r_i\}_{1 \le i \le n+1}, r_i = 0 \forall i = \overline{1,n}, r_{n+1} = 1.0$. Từ ma trận trọng số Z_Q chúng ta xác định vector thứ hạng $r_{cnn} = r^*_{CNN.Q,i}$bằng thuật toán EMR [42]

Bước 3: Lặp lại bước 2.1 - 2.3 nhưng với đặc trưng mức thấp $lf.E_i$ ta được giá trị xếp hạng là $r_{lf} = r^*_{lf.v.Q,i}$

Bước 4: **Kết hợp vector xếp hạng của 2 EMR:**

- Kết hợp tuyến tính $r^* = \{CB(r_{cnn,i}, r_{lf,i})\}_{1 \le i \le n}$, ở đó $CB(a,b) = \alpha * a + \beta * b$.

Kết quả trả về: r^*.

Đánh giá độ phức tạp của thuật toán CoEMR-S

Thuật toán **CoEMR-L** tương tự như thuật toán **CoEMR-S** nên thuật toán **CoEMR-L** có độ phức tạp tương đương với độ phức tạp của EMR gốc [42].

Nhận xét: Thuật toán CoEMR-L cung cấp một cách kết hợp tuyến tính xếp hạng đặc trưng mức thấp và xếp hạng cho đặc trưng CNN một cách hiệu quả, làm đa dạng đặc trưng và làm giàu thông tin từ các đặc trưng ảnh từ đó làm tăng độ chính xác của kết quả truy vấn hình ảnh trong CBIR.

2.5 Kết hợp luỹ thừa bậc lẻ 2 xếp hạng EMR.

Ý tưởng đề xuất kết hợp luỹ thừa bậc lẻ xuất phát từ việc nhận thấy rằng mỗi loại đặc trưng ảnh đều mang thông tin khác nhau: đặc trưng mức thấp thường liên quan đến các đặc điểm cơ bản của ảnh như cạnh, góc, điểm nổi bật, trong khi đặc trưng CNN (ví dụ: EfficientNet) có khả năng học các đặc trưng trừu tượng và phức tạp hơn, như các đối tượng, đặc điểm biên, hoặc mẫu hình học [108-109]. Luận án tìm cách đưa ra một mô hình kết hợp luỹ thừa bậc lẻ để có thể tận dụng được tính chất của toàn bộ các đặc trưng, khi đó lớp hàm

xếp hạng mới có thêm các tính chất như:

- Tính tổng hợp thông tin đa mức: Kết hợp xếp hạng luỹ thừa bậc lẻ giúp mô hình có khả năng tổng hợp thông tin đa mức, từ đó cải thiện khả năng nhận diện và hiểu bản chất của hình ảnh. Đặc trưng mức thấp giúp mô hình nhận diện các đặc điểm cơ bản của ảnh, trong khi đặc trưng CNN giúp mô hình học và phát hiện các đặc trưng phức tạp và trừu tượng

- Tính đa dạng của đặc trưng: Kết hợp xếp hạng luỹ thừa bậc lẻ giúp tăng tính đa dạng của các đặc trưng được sử dụng trong mô hình. Điều này có thể đồng nghĩa với việc mô hình có khả năng học và phát hiện nhiều loại đặc trưng khác nhau từ dữ liệu ảnh, từ các đặc điểm cơ bản đến các đặc trưng phức tạp, mô tả cả tính cục bộ và tính toàn cục của ảnh, giúp cải thiện khả năng tổng quát hóa và độ chính xác của mô hình. Phương pháp này sẽ được phân tích chi tiết trong các phần tiếp theo của luận án, bao gồm các bước thực hiện cụ thể và các kết quả thực nghiệm minh họa cho hiệu quả của phương pháp.

Trong phần này luận án đề xuất kỹ thuật kết hợp luỹ thừa bậc lẻ 2 bộ xếp hạng EMR cho đặc trưng mức thấp, đặc trưng CNN và xây dựng mô hình tra cứu ảnh dựa vào thuật toán CoEMR-P.

Kết hợp CoEMR kiểu luỹ thừa bậc lẻ (CoEMR-P)

Giả sử rằng trong tập dữ liệu ảnh gốc, mỗi ảnh có một vector đặc trưng duy nhất tương ứng và với một ảnh truy vấn I_Q, hệ thống chọn xếp hạng theo hình ảnh đặc trưng mức thấp $r^*_{lf.v.Q,i}$ kết hợp với xếp hạng của hình ảnh đặc trưng CNN $r^*_{CNN.Q,i}$, trong đó các vector xếp hạng được tính bằng EMR.

Sau đây là thuật toán chi tiết mà luận án đề xuất.

Thuật toán 2.3. CoEMR-P (Thuật toán kết hợp luỹ thừa bậc lẻ 2 bộ xếp hạng EMR).

Input: $\{I_i\}_{1\le i\le n}$ là tập ảnh gốc, I_Q ảnh truy vấn.

MxN là kích thước ảnh chuẩn hóa cho huấn luyện theo thuật toán

EfficientNet.

d: số chiều của một vector đặc trưng ảnh.

C: số lượng điểm neo của thuật toán EMR, tham số a ∈ (0,1) (a ≈ 1),

Output: r^* là thứ hạng tương tự với ảnh I_Q của ảnh I_i trong CSDL ảnh E.

Bước 1 (ngoại tuyến): **Xây dựng đồ thị EMR**

Bước 1a. Huấn luyện tập ảnh E theo mô hình EfficientNet với kích thước chuẩn hóa ảnh đầu vào là M*N, thu nhận được:

1.1. Tham số mô hình W_{CNN}.

1.2. Chạy từng ảnh $\{I_i\}_{1\le i\le n}$ qua mô hình thu được tập các vector đặc trưng $\{v.E_i\}_{1\le i\le n}$ với chiều d_{CNN}=1280.

1.3. Xác định C điểm neo $\{A_c\}_{1\le c\le C}$ của tập ảnh $\{I_i\}_{1\le i\le n}$ theo thuật toán K-means, sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$

1.4. Xác định ma trận kề $W=(w_{ij})$ của thuật toán EMR sử dụng các vector đặc trưng CNN $\{v.E_i\}_{1\le i\le n}$.

1.5. Xác định ma trận trọng số Z kích thước C*n của EMR.

Bước 1b. Trích rút đặc trưng mức thấp: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST, thu được tập véc tơ đặc trưng với số chiều d_{LF}=809. Lặp lại bước 1.3 - 1.5 như ở bước 1a với đặc trưng mức thấp.

Bước 2 (trực tuyến): Kết hợp xếp hạng 2 EMR

2.1. Chuẩn hóa ảnh I_Q về kích thước M*N, chạy qua mô hình EfficientNet đã xác định ở bước 1.1, thu nhận được vector đặc trưng CNN $v.E_Q$ với d chiều.

2.2. Mở rộng ma trận Z theo EMR [42]: Sử dụng C giá trị khoảng cách của E_Q với các phần tử neo của A_c, chúng ta thu được ma trận trọng số Z_Q mới có kích thước C*(n+1).

2.3. Đặt $r_Q = \{r_i\}_{1\le i\le n+1}, r_i = 0 \forall i = \overline{1,n}, r_{n+1} = 1.0$. Từ ma trận trọng số Z_Q chúng ta xác định vector thứ hạng $r_{cnn} = r^*_{CNN.Q,i}$ bằng thuật toán EMR [42].

Bước 3: Lặp lại bước 2.1 - 2.3 nhưng với đặc trưng mức thấp lf.E_i ta được giá trị xếp hạng là $r_{lf} = r^*_{lf.v.Q,i}$

Bước 4: **Kết hợp vector xếp hạng của 2 EMR:**

$r^* = \{CB(r_{cnn,i}, r_{lf,i})\}_{1 \le i \le n}$, ở đó $CB(a,b)$ cho bởi:

$$CB(a,b) = \begin{cases} \sqrt[3]{\frac{\{a^3 + b^3\}}{2}} & if\, a + b \ge 0 \\ \min(a,b) \end{cases}$$

Kết quả trả về: r^*.

Đánh giá độ phức tạp của thuật toán CoEMR-P

Thuật toán **CoEMR-P** có độ phức tạp tương đương với độ phức tạp của EMR gốc [42].

Nhận xét: Thuật toán CoEMR-P cung cấp một cách kết hợp tuyến tính xếp hạng đặc trưng mức thấp và xếp hạng cho đặc trưng CNN một cách hiệu quả, làm đa dạng đặc trưng và làm giàu thông tin từ các đặc trưng ảnh từ đó làm tăng độ chính xác của kết quả truy vấn hình ảnh trong CBIR.

Ba phương pháp **CoEMR-S, CoEMR-L và CoEMR-P** được đề xuất riêng rẽ thay vì gộp thành một phương pháp duy nhất nhằm đảm bảo tính linh hoạt trong xử lý dữ liệu và tối ưu hóa hiệu suất tra cứu ảnh cho từng trường hợp cụ thể. Mỗi phương pháp có cơ chế kết hợp riêng biệt, giúp khai thác hiệu quả ưu điểm của từng bộ xếp hạng EMR trên đặc trưng mức thấp và mức cao.

Cụ thể, **CoEMR-S** phù hợp khi một bộ xếp hạng có độ tin cậy cao hơn đáng kể so với bộ còn lại, cho phép lựa chọn phương án tốt nhất dựa trên một ngưỡng xác định. Phương pháp này đơn giản, hiệu quả và giảm thiểu tác động của bộ xếp hạng kém tin cậy. Trong khi đó, **CoEMR-L** sử dụng phương pháp kết hợp tuyến tính, cho phép điều chỉnh mức độ ảnh hưởng của từng bộ xếp hạng thông qua trọng số, giúp cân bằng thông tin từ cả hai nguồn đặc trưng. Phương pháp này linh hoạt và phù hợp với các trường hợp cả hai bộ xếp hạng

đều đóng góp quan trọng vào kết quả cuối cùng. **CoEMR-P**, với cơ chế kết hợp phi tuyến dựa trên lũy thừa bậc lẻ, giúp khuếch đại sự khác biệt giữa hai bộ xếp hạng và tăng cường khả năng phân biệt khi dữ liệu có quan hệ phi tuyến mạnh.

Việc đề xuất ba phương pháp tách biệt giúp đơn giản hóa quá trình tối ưu tham số, tránh làm giảm hiệu quả của từng phương pháp khi áp dụng chung một mô hình tổng quát. Đồng thời, điều này cũng tạo điều kiện thuận lợi cho việc so sánh, đánh giá hiệu suất trên nhiều bộ dữ liệu khác nhau, cũng như mở rộng nghiên cứu trong tương lai. Mỗi phương pháp có thể được điều chỉnh hoặc cải tiến riêng biệt tùy theo đặc điểm của tập dữ liệu và yêu cầu cụ thể của bài toán tra cứu ảnh.

2.6. Thực nghiệm và đánh giá kết quả

Các thực nghiệm được thực hiện trên máy tính cá nhân với cấu hình: CPU Intel® Core™ i5-6850HQ CPU @2.4 Ghz; GPU 4Ghz, RAM 32GB DDR3, Disk: 1000GB.

2.6.1 Đánh giá hiệu quả của của thuật toán CoEMR

Để đánh giá hiệu quả của thuật toán CoEMR, luận án tiến hành thực nghiệm tổ hợp các xếp hạng EMR cho đặc trưng mức thấp, đặc trưng mức cao trên các bộ dữ liệu theo Bảng 2.1 cụ thể như sau:

Bảng 2.1: Bảng kết quả các phương pháp tổ hợp xếp hạng EMR

TT	Bộ dữ liệu	Bộ đặc trưng 1	Bộ đặc trưng 2	Phương pháp tổ hợp	Kết quả công bố
1	VGG60K	Mức thấp	Véc tơ nhúng	Tuyến tính (CoEMR-L)	CT1, CT2
2	SIGN300HD	Mức thấp	Véc tơ nhúng	Lựa chọn (CoEMR-S)	CT4
3	Leaf30	Mức thấp	CNN	Tuyến tính (CoEMR-P)	CT6

Trong các thử nghiệm này, luận án sử dụng đặc trưng mức thấp (LF) gồm 5 bộ: Color Moments, LBP, Gabor Wavelets Texture, Edge và GIST để mô tả một hình ảnh [CT1-CT3]. Tất cả các đặc trưng này của các bộ dữ liệu Leaf30, Corel30k, VGG60K được chuẩn hóa sao cho mỗi thành phần vector của mỗi ảnh nằm trong phạm vi [-1, 1] và sau đó được nối thành một vetor với kích thước $d_{LF}=809$ [36].

Song song với nó, mỗi hình ảnh trong bộ dữ liệu này được chuẩn hóa về kích thước 256x256 và được đưa vào mô hình EfficentNet (đã cắt bỏ lớp cuối cùng) thu được tập vector đặc trưng CNN tương ứng có kích thước $d_{CNN}=1280$.

Trong thực nghiệm với bộ dữ liệu SIGN300HD, luận án sử dụng hai (02) đặc trưng mức thấp (LF) gồm: LBP, HOG để mô tả hình ảnh chữ ký, được chuẩn hoá thành một véc tơ với số chiều $d_{LF}=779$ [CT5].

Song song với nó, mỗi hình ảnh trong bộ dữ liệu này được chuẩn hóa về kích thước 256x256 và được đưa vào mô hình SigNet, mô hình này sử dụng Contrastive loss và một véc tơ nhúng $d_{eb}=128$.

Trong thực nghiệm này, luận án sử dụng *nbest = 350* cho phần phân cụm; Đối với thuật toán CoEMR, luận án chọn $s=5$ (s điểm neo láng giềng gần nhất của một vector đặc trưng ảnh).

Tất cả các tham số thực nghiệm cho từng bộ dữ liệu được thể hiện ở bảng sau:

Bảng 2.2: Bảng tham số thực nghiệm tra cứu ảnh sử dụng CoEMR

TT	Tập dữ liệu	N_{best}	C	r	α	β
1	Leaf30	16	200	3	0.8	0.2
2	Sign300HD	100	2000	5	0.2	0.8
3	VGG60K	120	5000	5	0.3	0.7

Với N_{best} là số ảnh cho mỗi lần truy vấn trong 20% số lượng mẫu truy vấn

lấy ngẫu nhiên, C là số điểm neo, C được chọn tùy theo bộ dữ liệu, r là số lân cận; α, β là hệ số sử dụng trong thuật toán CoEMR-L.

Để đánh giá khách quan hiệu quả của thuật toán EMR gốc và CoEMR đề xuất trên cùng các tập dữ liệu, luận án sử dụng một chỉ số tương tự độ đo Average Precision (ký hiệu AP) được đề xuất bởi NISTTREC video (TRECVID) [97, CT1], *AP* được định nghĩa trung bình của giá trị độ chính xác thu được sau mỗi ảnh liên quan được tra cứu.

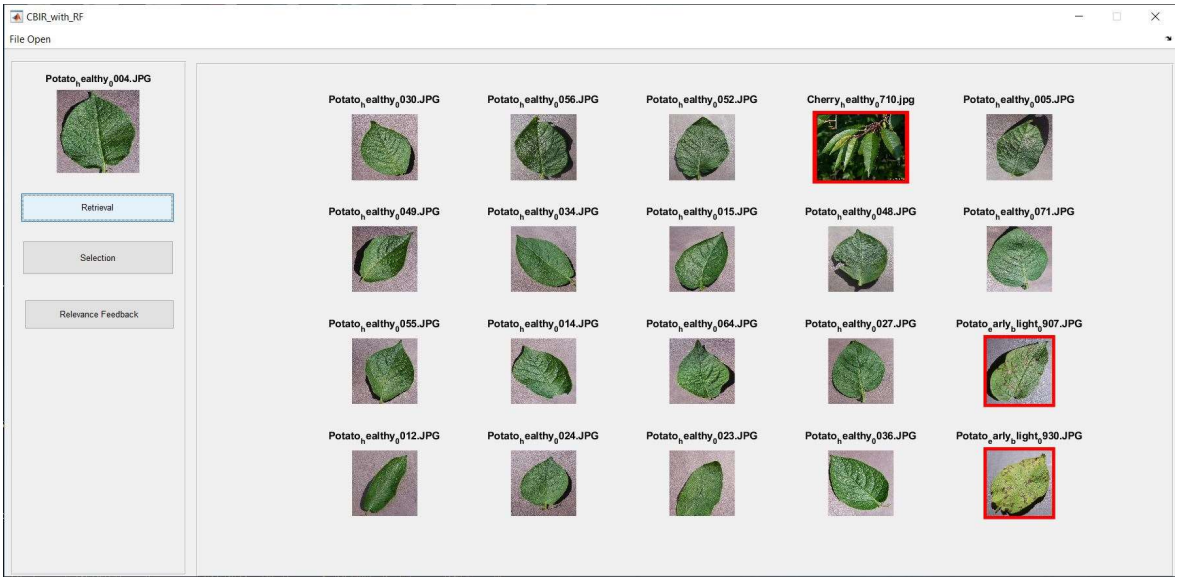
Tập ảnh truy vấn Q được chọn ngẫu nhiên với số lượng 20% ảnh từ mỗi lớp theo từng chủ đề của tập ảnh thử nghiệm Leaf30, Corel30K và VGG60K.

Đối với mỗi hình ảnh truy vấn $q \in Q$, sử dụng số liệu tương tự do EMR đưa ra, chọn N = 100 là số lượng hình ảnh trong một lớp. Giá trị chính xác là tỷ lệ trung bình giữa số lượng hình ảnh có liên quan N_q^+ trong hình ảnh N được tra cứu bởi sự giống nhau của từng hình ảnh q. Gọi tập hợp các phần tử có liên quan đến truy vấn $q \in Q$ là $\{d_1, d_2, \ldots, d_{mj}\}$, khi đó mAP(q) là độ chính xác cho một truy vấn q và mAP là độ chính xác cho tất cả các truy vấn được tính như sau: $mAP_{(q)} = \frac{N_q^+}{N} * 100$ (2.4)

và $mAP = \frac{1}{|Q|} \sum_{q=1}^{|Q|} mAP_{(q)}$ (2.5)

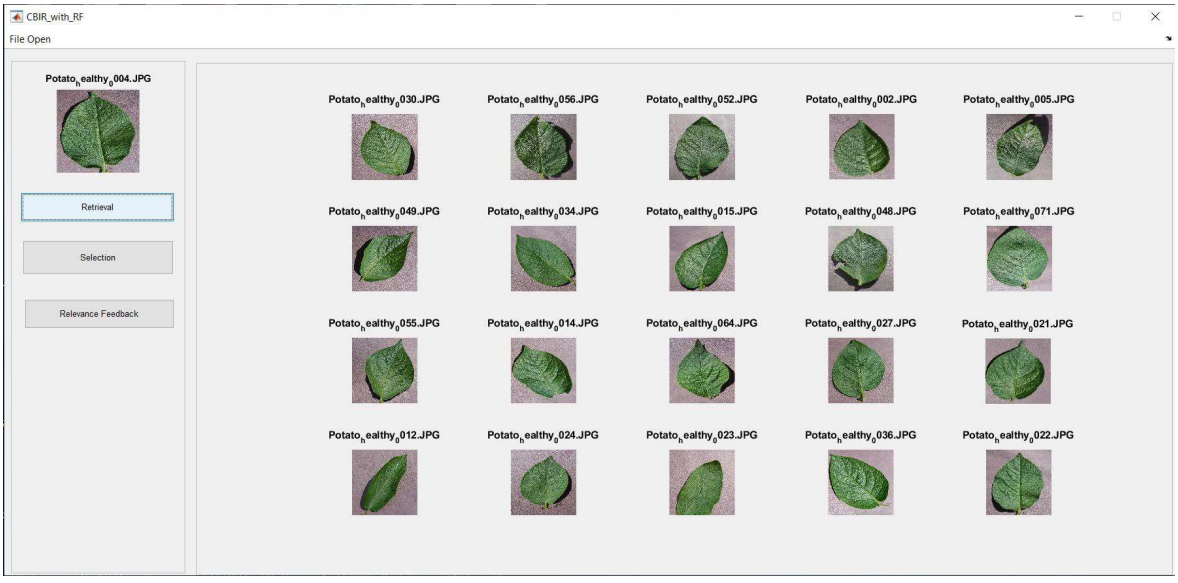
Các kết quả xếp hạng của EMR theo từng ảnh truy vấn q có thể xem như là một độ đo tương tự ảnh với mức độ tương tự giữa ảnh q và ảnh E_i trong CSDL được gán là $rank_q(E_i)$

Trong hình 2.4 là kết quả tra cứu đối với ảnh Potato_healthy_004.jpg trong tập Leaf30 với thuật toán EMR gốc với số lượng điểm neo là 2000. Kết quả tra cứu trả về là 20 ảnh có thứ hạng tốt nhất, kết quả có 03 ảnh (không liên quan so với ảnh truy vấn).



Hình 2.4. Kết quả truy vấn của ảnh Potato_healthy_004.jpg trong tập Leaf30 với thuật toán EMR gốc có 5 ảnh sai.

Trong hình 2.4, ảnh truy vấn Potato_healthy_004.jpg trong tập Leaf30 với thuật toán CoEMR với số lượng điểm neo là 2000, kết quả tra cứu trả về kết quả là 20 ảnh có thứ hạng tốt nhất đều liên quan (độ chính xác tra cứu khi kết quả trả về là 20 ảnh đạt 100% của ảnh này).



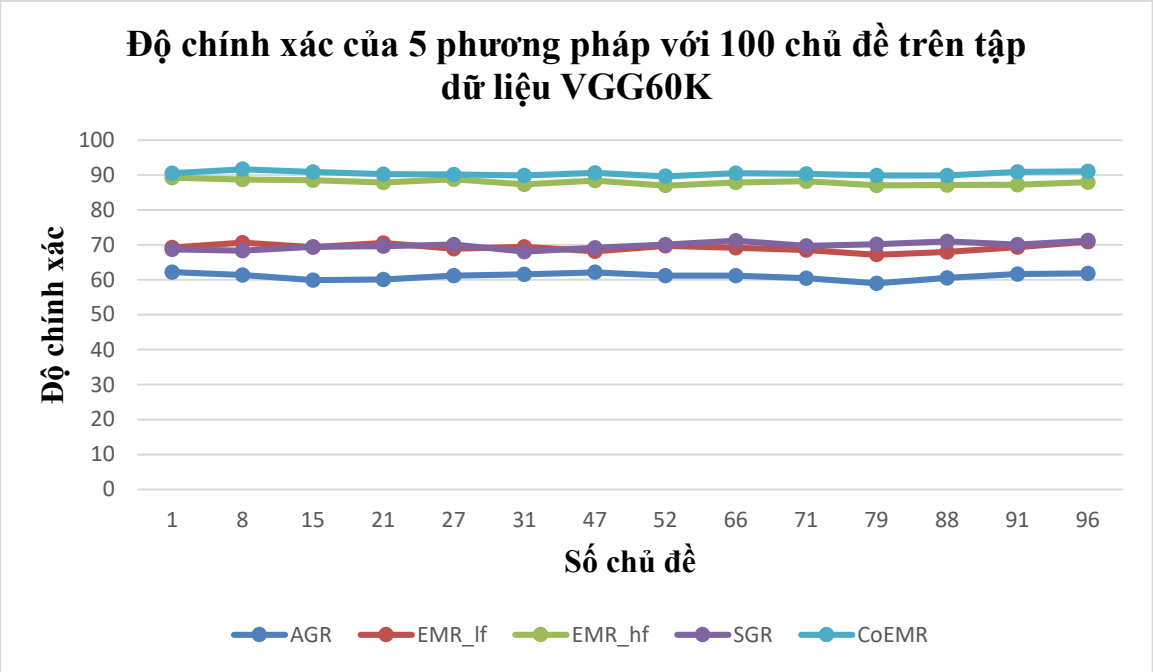
Hình 2.5. Kết quả truy vấn của ảnh Potato_healthy_004.jpg trong tập Leaf30 với thuật toán CoEMR.

2.6.2 Kết quả thực nghiệm

Để đánh giá của phương pháp đề xuất, luận án thực hiện so sánh kết quả khi sử dụng EMR cho xếp hạng đặc trưng mức thấp, sử dụng EMR cho xếp hạng đặc trưng CNN và sử dụng kết hợp xếp hạng theo thuật toán đề xuất CoEMR.

Ngoài ra luận án còn so sánh với các phương pháp xếp hạng dựa trên đồ thị khác, cụ thể: Anchor Graph Regularization (AGR) [38], Efficient Manifold Ranking (EMR) [42], Sub-graph regularization (SGR) [39] (là các kỹ thuật lựa chọn điểm neo dựa vào thuật toán K-means) trên các tập dữ liệu VGG60K, Leaf30, Sign300HD, Corel30K. Lý do so sánh thuật toán CoEMR với AGR, SGR và EMR là vì các thuật toán đều cải tiến thuật toán MR và sử dụng thuật toán phân cụm Kmeans để tìm điểm neo.

Để tiến hành thực nghiệm, luận án chọn ngẫu nhiên 100 chủ đề (trong tổng số 500 chủ đề của tập dữ liệu gốc, mỗi chủ đề có 120 ảnh) trên tập dữ liệu VGG60K và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về 20 ảnh sau mỗi lần tra cứu. Từ kết quả hình 2.5 ta thấy phương pháp AGR cho độ chính xác thấp nhất. Cũng trong hình này, độ chính xác trung bình của phương pháp đề xuất là cao nhất so với các phương pháp còn lại.

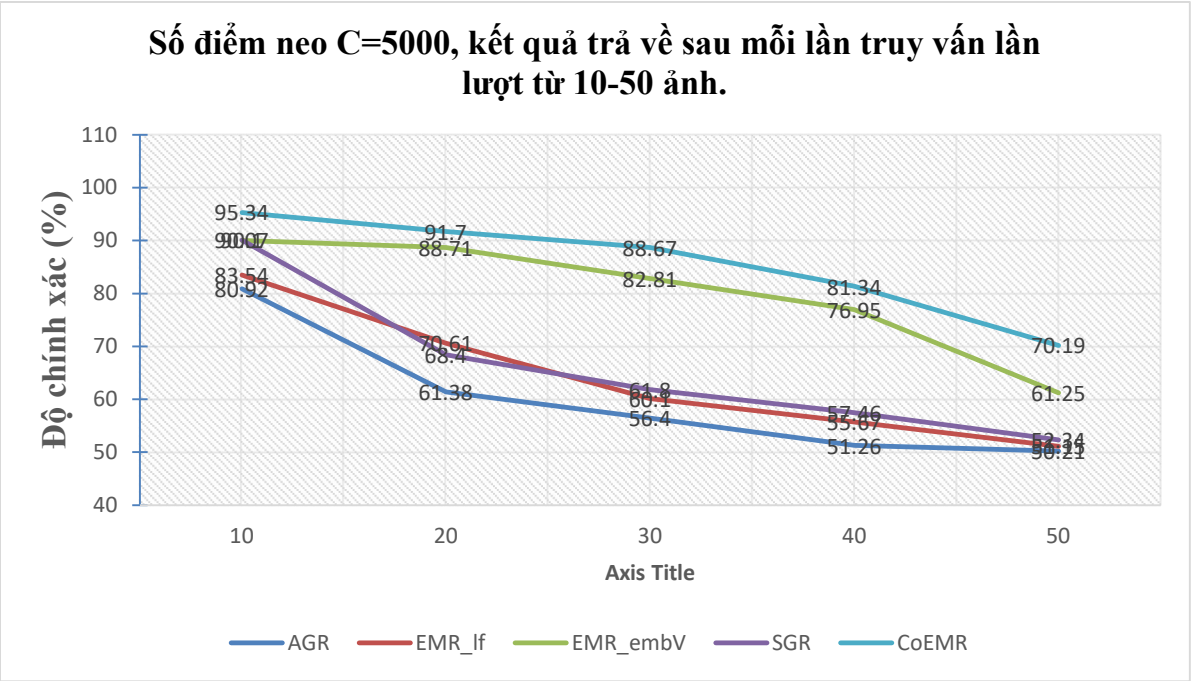


Hình 2.6. Độ chính xác của 5 phương pháp với 100 chủ đề trên tập dữ liệu VGG60K.

Trong các thực nghiệm tiếp theo, luận án thực nghiệm với phạm vi trên các tập dữ liệu VGG60K, Leaf30, Sign300HD, số ảnh được xếp hạng cao nhất lần lượt trả về sau tra cứu là 10, 20, 30, 40, 50 của năm phương pháp. Với số lượng điểm neo được lựa chọn là theo từng bộ dữ liệu tại Bảng 2.4 . Các đường cong trung bình độ chính xác - phạm vi (average precision-scope curves) được thể hiện trong hình 2.6 trên tập dữ liệu VGG60K. Với phạm vi số ảnh trả về 10, 20, 30, 40, 50 phương pháp đề xuất cho hiệu quả cao hơn các phương pháp còn lại từ 6% đến 15%.

Đối với thực nghiệm trên tập dữ liệu VGG60K, luận án chọn ngẫu nhiên 100 chủ đề (trong tổng số 500 chủ đề của tập dữ liệu gốc VGG , mỗi chủ đề có 120 ảnh) và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về sau tra cứu lần lượt là 10, 20, 30, 40, 50 ảnh. Các tham số được đặt chung cho tất cả các thực nghiệm với bộ dữ liệu VGG60K như sau: tham số a = 0.99, số điểm neo C = 5000, r = 5, nbest cho mỗi lần truy vấn n_b= 120, 20% số lượng

mẫu truy vấn lấy ngẫu nhiên. Tham số kết hợp 2 xếp hạng trong CoEMR là α = 0.3, β=0.7.



Hình 2.7. Độ chính xác trung bình tra cứu trên tập ảnh VGG60K.

Số liệu độ chính xác trung bình kết quả tra cứu trên tập ảnh VGG60k tính theo MAP được thể hiện ở bảng 2.3.

Bảng 2.3: Độ chính xác của 5 phương pháp ở 20 ảnh trả về sau tra cứu trên tập dữ liệu VGG60K

PP tra cứu / Số ảnh trả về	AGR (%)	EMR_{LF} (%)	EMR_{EmbV} (%)	SGR (%)	CoEMR (%)
10	80.92	83.54	90.07	90.10	**95.34**
20	61.38	70.61	88.71	68.40	**91.70**
30	56.40	60.10	82.81	61.80	**88.67**
40	51.26	55.67	76.95	57.46	81.34
50	50.21	51.15	61.25	52.34	70.19

Đối với thực nghiệm trên tập dữ liệu Leaf30, luận án chọn toàn bộ 58 chủ đề, mỗi chủ đề chọn ngẫu nhiên 100 ảnh và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về sau tra cứu lần lượt là 10, 20, 30 ảnh. Các tham số được đặt chung cho tất cả các thực nghiệm với bộ dữ liệu Leaf30 như sau: tham số a = 0.99, số điểm neo C = 2000, r = 5, nbest cho mỗi lần truy vấn n_b= 100, 20% số lượng mẫu truy vấn lấy ngẫu nhiên. Tham số kết hợp 2 xếp hạng trong CoEMR là α = 0.2, β=0.8. Số liệu độ chính xác trung bình kết quả tra cứu trên tập ảnh Leaf30 tính theo MAP được thể hiện ở bảng 2.4.

Bảng 2.4: Độ chính xác của 5 phương pháp ở 20 ảnh trả về sau tra cứu trên tập dữ liệu Leaf30

PP tra cứu / Số ảnh trả về	AGR (%)	EMR_{LF} (%)	EMR_{CNN} (%)	SGR (%)	CoEMR (%)
10	59.96	68.12	75.67	73.89	**83.42**
20	57.64	65.05	70.72	69.23	**81.70**
30	52.34	61.26	68.67	66.27	**80.33**

Đối với thực nghiệm trên tập dữ liệu Sign300HD, luận án chọn 100 chủ đề trong 300 chủ đề, mỗi chủ đề có 16 ảnh và thực hiện chọn ảnh truy vấn trong tất cả các chủ đề. Kết quả trả về sau tra cứu lần lượt là 10, 13, 16 ảnh. Các tham số được đặt chung cho tất cả các thực nghiệm với bộ dữ liệu Sign300HD như sau: tham số a = 0.99, số điểm neo C = 500, r = 3, nbest cho mỗi lần truy vấn n_b= 16, 20% số lượng mẫu truy vấn lấy ngẫu nhiên. Tham số kết hợp 2 xếp hạng trong CoEMR là α = 0.8, β=0.2. Số liệu độ chính xác trung bình kết quả tra cứu trên tập ảnh Sign300HD tính theo MAP được thể hiện ở bảng 2.5.

Bảng 2.5: Độ chính xác của 4 phương pháp ở 16 ảnh trả về sau tra cứu trên tập dữ liệu Sign300HD

PP tra cứu / Số ảnh trả về	AGR (%)	EMR_{LF} (%)	EMR_{SigNet} (%)	SGR (%)	CoEMR (%)
10	45.67	52.13	54.16	51.15	**64.46**
13	42.33	49.51	52.16	48.81	**61.36**
16	40.16	42.89	45.67	42.26	**56.67**

Từ kết quả ở bảng 2.3, 2.4, 2.5 ta thấy phương pháp AGR cho độ chính xác thấp nhất. Cách lựa chọ điểm neo trong AGR là chọn ngẫu nhiên tập điểm đại diện trong CSDL làm điểm neo. Đối với 2 phương pháp SGR và EMR, phương pháp lựa chọn điểm neo là dùng thuật toán phân cụm K-means. Cũng trong bảng này, độ chính xác trung bình của phương pháp đề xuất là cao nhất so với các phương pháp còn lại.

2.7. Kết hợp nhiều truy vấn ảnh trong CBIR

Trong hầu hết các hệ thống CBIR, mỗi truy vấn hình ảnh thường chỉ sử dụng một ảnh duy nhất, dẫn đến việc thiếu thông tin và hạn chế trong việc mô tả đầy đủ ý nghĩa ngữ nghĩa của ảnh. Điều này có thể làm giảm độ chính xác của kết quả truy vấn. Một hướng nghiên cứu tiềm năng để giải quyết vấn đề này là sử dụng truy vấn nhiều ảnh, nhằm cải thiện hiệu quả truy vấn. Việc sử dụng đồng thời nhiều truy vấn giúp khai thác nhiều thông tin hơn từ cơ sở dữ liệu và cung cấp kết quả tìm kiếm chính xác hơn [141-143]. Tuy nhiên, việc kết hợp kết quả từ nhiều truy vấn vẫn là một thách thức lớn.

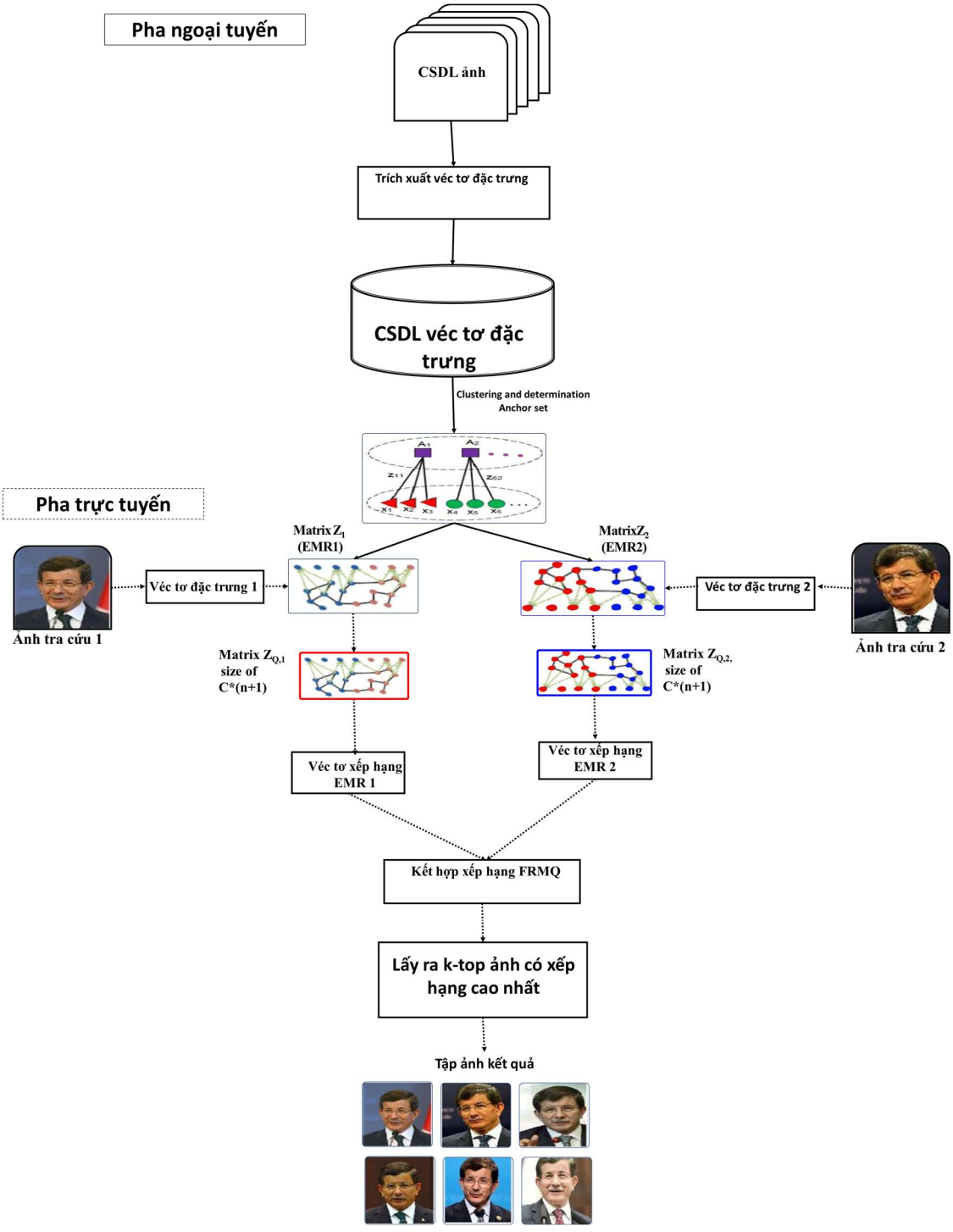
Sử dụng nhiều truy vấn (đa truy vấn) là một phương pháp tiếp cận hiệu quả trong lĩnh vực tra cứu thông tin và tìm kiếm dữ liệu, nổi bật với khả năng sử dụng đồng thời nhiều truy vấn để cải thiện chất lượng của kết quả truy vấn

[143]. Lợi ích chính của đa truy vấn là tăng độ chính xác và độ tin cậy của thông tin được trả về. Thay vì chỉ dựa trên một truy vấn duy nhất, đa truy vấn cho phép hệ thống tra cứu dữ liệu từ nhiều góc độ khác nhau, từ đó đưa ra kết quả tổng thể và chi tiết hơn về nhu cầu tìm kiếm của người dùng[141].

Việc sử dụng nhiều truy vấn giúp giảm thiểu sai sót trong quá trình tra cứu dữ liệu [141-143]. Mỗi truy vấn đơn lẻ có thể bỏ qua một số khía cạnh quan trọng của dữ liệu, nhưng khi kết hợp nhiều truy vấn, hệ thống có thể bao quát hơn và đáp ứng đầy đủ các yêu cầu của người dùng. Đặc biệt, việc phản ánh tốt hơn ý định người dùng là một điểm mạnh của đa truy vấn, bởi người dùng thường có nhiều mục đích và khía cạnh khác nhau khi tìm kiếm thông tin [144].

Như vậy cần đảm bảo cân bằng giữa các truy vấn. Đảm bảo rằng mỗi truy vấn được xem xét công bằng và không bị lấn át bởi truy vấn khác là điều cần thiết để đảm bảo tính công bằng và đúng đắn trong quá trình tra cứu dữ liệu [143]. Do đó, phát triển và triển khai đa truy vấn đòi hỏi sự kết hợp của các thuật toán thông minh và sự hiểu biết sâu sắc về cấu trúc dữ liệu và yêu cầu của người dùng. Các phương pháp hiện tại hoặc gộp các truy vấn thành một truy vấn duy nhất hoặc xem chúng một cách độc lập. Điều này có thể dẫn đến mất thông tin quan trọng [143]. Nhiều phương pháp sử dụng đa truy vấn được đưa ra, như: Sử dụng học trọng số đặc trưng [141-142], Xếp hạng song song đa truy vấn (Multi-Query parallel field ranking for image retrieval) [143], tuy nhiên việc đưa ra một phương pháp hiệu quả để tổng hợp và xử lý các kết quả từ các truy vấn khác nhau là một vấn đề phức tạp, đặc biệt khi dữ liệu lớn và phức tạp.

Để cải thiện độ chính xác khi sử dụng đa truy vấn trong CBIR, luận án đề xuất phương pháp kết hợp 2 xếp hạng EMR cho 2 ảnh truy vấn gọi là Fusion Ranking Multi-Query (EMR-FRMQ) [CT8]. Hình 2.8 Mô tả sơ đồ mô hình CBIR sử dụng kết hợp 2 xếp hạng EMR cho 2 ảnh truy vấn.



Hình 2.8. Sơ đồ mô tả hệ thống (CBIR) sử dụng đa truy vấn.

Phương pháp này khai thác hiệu quả từ nhiều truy vấn và sự phân bố nội tại của dữ liệu, từ đó tăng cường hiệu quả tra cứu hình ảnh trong hệ thống CBIR một cách đáng kể [CT8], cụ thể bảng 2.6 mô tả kết quả sử dụng CBIR với đa truy vấn ảnh.

Bảng 2.6: Độ chính xác của phương pháp sử dụng đa truy vấn ở 20 ảnh trả về sau tra cứu trên tập dữ liệu VGG60K

Đặc trưng	Sử dụng 1 ảnh truy vấn	Sử dụng 2 ảnh truy vấn
Đặc trưng mức thấp	70.61%	**76.21%**
Đặc trưng CNN (EfficientNET)	87.15%	**90.46%**
Đặc trưng vector nhúng (FaceNet)	88.71%	**92.67%**

2.8 Kết luận Chương 2

Trong chương này, luận án đã trình bày chi tiết phương pháp đề xuất - kết hợp xếp hạng EMR cho đặc trưng mức thấp và mức cao (đặc trưng nhúng, đặc trưng CNN). Phương pháp này mang lại một hướng tiếp cận tổng hợp, giúp khai thác tối đa thông tin từ cả hai loại đặc trưng, khắc phục các hạn chế khi một trong hai loại đặc trưng bị suy giảm độ chính xác xếp hạng trong các cơ sở dữ liệu ảnh thực tế. Bên cạnh đó, trong chương này luận án cũng trình bày đề xuất cũng phương pháp kết hợp xếp hạng cho cả các hệ thống CBIR với nhiều ảnh truy vấn. Kết quả thực nghiệm cho thấy, việc tổng hợp kết quả xếp hạng EMR từ từng ảnh truy vấn giúp nâng cao đáng kể độ chính xác tra cứu ảnh so với hệ thống chỉ sử dụng một ảnh truy vấn.

Cụ thể, đặc trưng mức thấp có ưu điểm ít phụ thuộc vào độ phân giải ảnh nhưng lại nhạy cảm với nhiễu và chất lượng ảnh kém. Trong khi đó, đặc trưng mức cao (vector nhúng từ CNN) có khả năng phân biệt mạnh nhưng bị ảnh hưởng bởi độ phân giải và sự khác biệt trong phân bố thống kê của dữ liệu huấn luyện. Việc kết hợp tuyến tính và phi tuyến giữa hai bộ xếp hạng giúp tận dụng ưu điểm của cả hai loại đặc trưng, nâng cao độ chính xác trong truy vấn ảnh.

Luận án đã đề xuất và đánh giá thực nghiệm các thuật toán CoEMR-L, CoEMR-S và CoEMR-P trên các bài toán tra cứu ảnh dựa trên nội dung (CBIR) với một ảnh truy vấn (CT1-CT4) và nhiều ảnh truy vấn (CT8). Kết quả thực nghiệm trên các tập dữ liệu Leaf30, Sign300HD, Corel30K và VGG60K cho thấy phương pháp kết hợp này mang lại mức cải thiện đáng kể về độ chính xác (tăng khoảng 7% - 15% so với EMR gốc và các phương pháp khác cùng họ).

Luận án cũng đã mở rộng phương pháp EMR-FRMQ kết hợp tuyến tính và phi tuyến trên kết quả xếp hạng của từng ảnh riêng lẻ khi có nhiều ảnh truy vấn. Kết quả cho thấy việc tổng hợp thông tin từ nhiều ảnh truy vấn giúp hệ thống tra cứu ảnh hoạt động ổn định và chính xác hơn, đặc biệt trong các tình huống có độ nhiễu cao hoặc dữ liệu phức tạp.

Thuật toán đề xuất đã được đánh giá qua các chỉ số mAP và minP, cho thấy hiệu suất ổn định trên nhiều tập dữ liệu khác nhau. So sánh với các phương pháp hiện có như AGR, SGR và EMR, kết quả khẳng định CoEMR có độ chính xác cao hơn và độ ổn định tốt hơn, góp phần cải thiện đáng kể hiệu quả của hệ thống tra cứu ảnh.

Ngoài ứng dụng trong CBIR, phương pháp kết hợp xếp hạng EMR còn góp phần bổ sung vào hướng nghiên trong các lĩnh vực khác như nhận dạng đối tượng, phân loại ảnh và các hệ thống trí tuệ nhân tạo. Điều này không chỉ có giá trị về mặt lý thuyết mà còn có ý nghĩa thực tiễn cao trong việc phát triển các hệ thống tra cứu ảnh thông minh.

Thuật toán kết hợp xếp hạng EMR giữa các bộ đặc trưng mức thấp và mức cao đã chứng minh được hiệu quả trong việc nâng cao độ chính xác tra cứu ảnh, với mức cải thiện từ 7% - 15% so với EMR truyền thống. Đồng thời, luận án cũng mở rộng mô hình xếp hạng cho nhiều ảnh truy vấn, giúp tăng tính ứng dụng thực tế của hệ thống CBIR. Tuy rất mạnh mẽ, EMR vẫn tồn tại những hạn chế, đặc biệt trong việc so sánh độ tương tự giữa các ảnh nằm ngoài cơ sở dữ liệu. Do EMR hoạt động dựa trên cấu trúc đồ thị của tập dữ liệu có sẵn, việc

mở rộng ra các ảnh chưa được đưa vào CSDL trở nên khó khăn, làm giảm khả năng tổng quát hóa của mô hình. Hạn chế này gây trở ngại trong nhiều bài toán thực tế như nhận diện khuôn mặt, so khớp chữ ký và các ứng dụng bảo mật, nơi việc đánh giá độ tương tự giữa các ảnh chưa từng xuất hiện trong hệ thống là rất quan trọng.

Để giải quyết vấn đề này luận án cần xây dựng một độ đo tương tự ảnh kế thừa ưu điểm của thuật toán xếp hạng EMR (giá trị xếp hạng sự tương tự cao giữa ảnh truy vấn và các ảnh trong CSDL ảnh được lan truyền theo đồ thị quan hệ kề giữa các ảnh theo cấu trúc k-NN). Thuật toán xây dựng độ tương tự sẽ học các giá trị xếp hạng EMR giữa các cặp ảnh được chọn phù hợp của CSDL ảnh đã cho ở đó một ảnh xem như là ảnh truy vấn, mô hình học khái quát để đưa ra một độ đo tương tự ảnh giữa các cặp ảnh có thể hoàn toàn chưa được bổ sung vào CSDL ảnh đã cho. Chi tiết về cách xây dựng độ đo tương tự ảnh dựa trên các giá trị xếp hạng EMR đề xuất của luận án được trình bày ở chương 3: Xây dựng độ đo tương tự ảnh theo các giá trị xếp hạng EMR.

CHƯƠNG 3: XÂY DỰNG ĐỘ ĐO TƯƠNG TỰ ẢNH THEO CÁC GIÁ TRỊ XẾP HẠNG EMR

Trong tra cứu ảnh dựa trên nội dung (CBIR), độ đo khoảng cách giữa các đối tượng đóng vai trò quan trọng trong việc xác định mức độ tương đồng giữa chúng. Thay vì sử dụng độ đo Euclid mặc định, phương pháp học độ đo khoảng cách nhằm mục tiêu xây dựng một độ đo phù hợp hơn dựa trên thông tin có trong tập dữ liệu.

Cụ thể, phương pháp học độ đo khoảng cách nhằm tìm ra một ma trận độ đo M sao cho khoảng cách theo ma trận M giữa các đối tượng cùng lớp càng nhỏ, trong khi khoảng cách giữa các đối tượng khác lớp càng lớn. Điều này tương ứng với mục tiêu tối ưu hóa độ phân biệt của các lớp trong không gian đặc trưng.

Phương pháp học độ đo khoảng cách là một khía cạnh quan trọng của học máy, đã và đang được nghiên cứu rộng rãi. Nó đóng vai trò then chốt trong nhiều bài toán như phân loại, xếp hạng, ghép cặp và tra cứu ảnh. Thông thường, các phương pháp này sử dụng các ràng buộc của các cặp hoặc ba đối tượng để học ma trận khoảng cách.

Học độ đo tương tự là phương pháp quan trọng trong lĩnh vực CBIR nhằm xây dựng không gian đặc trưng phù hợp cho bài toán. Đây là một tiếp cận dựa trên giả định rằng càng giống nhau (tương tự) thì độ tương tự càng cao và ngược lại. Các phương pháp học độ đo tương tự nhằm xác định biến đổi sao cho phản ánh đúng mối quan hệ tương tự giữa các vector đặc trưng trong tập dữ liệu, từ đó cải thiện hiệu quả phân loại và tra cứu ảnh.

EMR là một thuật toán phổ biến và mang lại hiệu quả cao trong lĩnh vực truy vấn hình ảnh dựa trên nội dung (CBIR) [36, 42, 127], EMR sử dụng phương trình lan truyền xếp hạng để tính toán vector xếp hạng cho mỗi hình ảnh truy vấn, trong đó thành phần biểu diễn mức độ tương tự giữa hình ảnh truy

vấn và hình ảnh thứ trong tập dữ liệu, ngoài ra, EMR còn được sử dụng để đo độ tương tự ngữ nghĩa giữa các đối tượng, như đã được trình bày trong một số nghiên cứu [126]. So với các phương pháp đo khoảng cách thông thường chỉ xét cặp đôi, EMR tính toán độ tương đồng dựa trên cả mối quan hệ lân cận và lan truyền xếp hạng trên toàn bộ đồ thị, phản ánh đầy đủ hơn mối quan hệ giữa các mẫu dữ liệu. Điều này cho thấy tiềm năng rộng lớn của EMR trong việc khai thác các mối liên hệ tiềm ẩn giữa các điểm dữ liệu. Hình 3.1 mô tả 3 hình ảnh thuộc cùng một nhãn trong bộ dữ liệu Corel30K, nếu so sánh bằng "cảm quan" thì khó tìm ra sự tương đồng trong 3 bức ảnh này mặc dù nó có sự tương đồng về mặt ngữ nghĩa được chụp tại cùng một địa điểm ở châu Phi.

16005.jpg 16006.jpg 16007.jpg

Hình 3.1: Ba hình ảnh thuộc bộ dữ liệu Corel30K và được phân loại là có sự tương đồng ngữ nghĩa.

Tuy nhiên, EMR vẫn còn tồn tại một số nhược điểm cần khắc phục. Đầu tiên, EMR chỉ có thể xác định độ tương đồng giữa các hình ảnh trong tập dữ liệu ban đầu, không thể ứng dụng cho các hình ảnh ngoài tập dữ liệu này. Thứ hai, mỗi khi có sự thay đổi hoặc bổ sung dữ liệu mới, EMR phải xây dựng lại toàn bộ đồ thị, dẫn đến việc tính toán phức tạp và tốn kém tài nguyên, đặc biệt là khi làm việc với hệ thống dữ liệu lớn, cụ thể:

- EMR chỉ có thể xác định độ tương đồng giữa các hình ảnh trong tập dữ liệu ban đầu, không thể ứng dụng cho các hình ảnh ngoài tập dữ liệu này.

Trong thực tế, đối với các hệ thống nhận diện hoặc so khớp hình ảnh khuôn mặt hay chữ ký, việc so sánh hai ảnh không nằm trong cơ sở dữ liệu ảnh là một công việc thường xuyên và quan trọng. Các hệ thống này không chỉ cần so sánh các ảnh đã được lưu trữ trước đó mà còn phải xử lý và so sánh các ảnh mới, chưa từng xuất hiện trong cơ sở dữ liệu.

Chẳng hạn, trong một hệ thống nhận diện khuôn mặt, khi một người mới đăng nhập hoặc thực hiện một giao dịch, hệ thống phải so khớp khuôn mặt của người đó với dữ liệu đã có để xác minh danh tính. Tương tự, trong hệ thống xác thực chữ ký, khi một chữ ký mới được cung cấp, hệ thống phải so sánh nó với các mẫu chữ ký đã lưu trữ để xác định tính hợp lệ.

Mỗi khi thay đổi hoặc bổ sung thêm dữ liệu mới, EMR phải xây dựng lại toàn bộ đồ thị, dẫn đến tính toán phức tạp, nhất là với hệ thống dữ liệu lớn.

Nhằm áp dụng EMR vào bài toán thực tế khi so sánh độ tương tự giữa hai ảnh, chúng ta có thể xem EMR như một độ đo tương tự. Từ đó, luận án đề xuất việc xây dựng mô hình EMR learning, sử dụng kỹ thuật học máy để học trực tiếp từ kết quả xếp hạng của EMR. Mô hình EMR learning có khả năng dự đoán độ tương tự cho các hình ảnh nằm ngoài tập huấn luyện, giúp nâng cao hiệu quả và độ chính xác trong việc tra cứu và nhận dạng ảnh.

Chương này sẽ trình bày chi tiết phương pháp đề xuất, bắt đầu từ việc xây dựng độ đo tương tự EMR, xây dựng tập huấn luyện IC, tập véc tơ đầu vào, đầu ra VDR và các phương pháp học máy hồi quy. Cuối cùng, các thử nghiệm và đánh giá sẽ được thực hiện để kiểm chứng tính khả thi và hiệu quả của phương pháp đề xuất, từ đó khẳng định tiềm năng ứng dụng của EMR learning trong thực tế.

3.1 Mô hình học xếp hạng EMR

Để xây dựng mô hình học xếp hạng EMR, đầu tiên từ CSDL chứa n ảnh E, sau khi trích chọn đặc trưng ảnh chúng ta thu được CSDL n véc tơ đặc trưng ảnh E (có thể không có nhãn) và xây dựng đồ thị EMR, mô hình EMR learning là một bộ (IC, VDR, S) được khái quát như sau:

(i) $IC \subset \text{ExE} = \{(v_1, v_2) | v_1, v_2 \in E\},\ IC \neq \emptyset.$ (3.1)

Tập IC nói chung cần thỏa mãn yêu cầu sau:

$$\#\text{IC} \ll (\#E)^2 ,$$

nghĩa là số phần tử của IC bé hơn rất nhiều so số cặp ảnh của CSDL ảnh.

(ii) VDR là tập huấn luyện cho mô hình với đầu vào là cặp vector thuộc IC, đầu ra là mức độ tương tự được ước lượng bằng EMR.

(iii) Độ đo tương tự

Độ đo S được xây dựng từ tập VDR bằng một thuật toán học máy, S đo độ tương tự giữa các vector đặc trưng ảnh nằm ngoài CSDL vector đặc trưng ảnh. Do S là học đầu ra của xếp hạng EMR nên độ đo tương tự S của mô hình EMR learning cần thỏa mãn yêu cầu sau:

$S(v_1, v_2) \approx \text{r}[index(v_2)]$ trong đó: $\vec{r}$=EMR(v_1), $v_1, v_2 \in$ IC. (3.2)

3.2 Phương pháp xác định tập IC

Tiếp theo, luận án đề xuất thuật toán lựa chọn tập cặp vector đặc trưng ảnh IC của mô hình EMR learning, cụ thể gồm các bước như sau:

Bước 1: $IC = \emptyset$ (IC rỗng).

Bước 2: Từ cơ sở dữ liệu véc tơ đặc trưng E, chọn ra tập K các vector đặc trưng ảnh ngẫu nhiên (K có thể chọn bằng [n/2]).

Bước 3: Với mỗi vector đặc trưng ảnh $v_1 \in K$, quét toàn bộ CSDL E lấy ra n_b (ví dụ n_b=20) vector $\{v_{2,i}\}_{1 \le i \le n_b}$ đặc trưng gần nhất với v_1 đo theo xếp hạng của EMR.

Bước 4: Bổ sung $\{v_{1,i}, v_{2,i}\}_{1 \le i \le n_b}$ vào IC.

Kết quả: Ta có tập IC có số cặp vector đặc trưng ảnh là n_b*K.

Thuật toán lựa chọn tập cặp vector đặc trưng ảnh IC của mô hình EMR learning như sau:

Thuật toán 3.1. IC-S (ImageCoupleSelect)

Input:

E: CSDL vector đặc trưng ảnh,

K: Số lượng vector đặc trưng ảnh ngẫu nhiên được chọn,

n_b: Số lượng vector đặc trưng gần nhất cần lấy ra cho mỗi vector đặc trưng ảnh trong K.

Output: IC: Tập hợp cặp vector đặc trưng ảnh.

Bước 1. Khởi tạo $IC = \emptyset$ (IC rỗng)

Bước 2. Chọn K vector ngẫu nhiên, $Kset = \{v_{i_k}\}_{1 \le k \le K}$ đặc trưng ảnh từ CSDL E.

Bước 3. Với mỗi vector đặc trưng ảnh $v_1 \in Kset$,

- Quét toàn bộ CSDL E

- Lấy ra n_b vector $\{v_{2,i}\}_{i \le 1 \le n_b}$ đặc trưng gần nhất với v_1 dựa trên xếp hạng của EMR.

Bước 4. Bổ sung các cặp$\{(v_{1,i}, v_{2,i})\}_{i \le 1 \le n_b}$ vào IC

return IC.

Độ phức tạp thuật toán:

Giả sử n là số lượng vector trong CSDL E:

Bước chọn ngẫu nhiên K vector từ E có độ phức tạp O(K). Với mỗi vector $v_1 \in K$, việc quét toàn bộ CSDL E để tìm n_b vector gần nhất có độ phức tạp $O(n)$.

- Vậy độ phức tạp của việc tìm n_b vector gần nhất cho K vector là $O(K * n)$.

- Việc bổ sung các cặp vector vào IC có độ phức tạp $O(K * n_b)$.

Tổng hợp lại, độ phức tạp của thuật toán là: $O(K * n + K * n_b)$.

Nếu K xấp xỉ $\frac{n}{2}$, thì độ phức tạp là $O(n^2)$.

3.3 Phương pháp xác định tập huấn luyện của EMR learning

Sau khi có tập IC với số cặp véc tơ đặc trưng ảnh là n_b*K, luận án tiến hành các bước tạo cặp tập đầu vào - đầu ra phù hợp với thuật toán học máy có nhãn.

Quá trình này bao gồm các bước cụ thể như sau:

Đầu tiên, với mỗi cặp véc tơ đặc trưng $(v_1, v_2) \in IC$ luận án xây dựng một mẫu đầu vào bằng cách tạo ra vector khác biệt X= $\left(\left|v_{1,i} - v_{2,i}\right|\right)_{1 \le i \le \mathrm{d} \quad (v_1)}$, đây là vector chứa các giá trị tuyệt đối của sự chênh lệch giữa các thành phần tương ứng của hai véc tơ đặc trưng

Tiếp theo, giá trị đầu ra được xác định là một số cụ thể, đó là thứ hạng tương tự: $s = r[index(v_2)]$, trong đó: $\vec{r} = \mathrm{EMR}(v_1)$. Điều này có nghĩa là sử dụng kết quả xếp hạng EMR của v_1 để xác định mức độ tương tự của v_2 $so\ với\ v_1$.

Quá trình này được lặp lại cho toàn bộ các cặp véc tơ trong IC, từ đó thu được tập đầu vào và đầu ra (X, s), mà luận án đặt tên là VDR (vector difference ranking).

Sau khi xây dựng tập VDR, chia tập này thành hai bộ dữ liệu để phục vụ quá trình huấn luyện và kiểm tra mô hình, cụ thể:

- Bộ huấn luyện: 80% dữ liệu được sử dụng để huấn luyện SVR.

- Bộ kiểm tra: 20% còn lại dùng để kiểm tra độ chính xác dự đoán của mô hình.

Quá trình này đảm bảo rằng mô hình SVR được huấn luyện và kiểm tra một cách khách quan, từ đó đánh giá được hiệu suất và độ chính xác của mô hình trong việc dự đoán thứ hạng tương tự của các ảnh.

Bước tiếp theo, luận án xây dựng thuật toán EMRTrainingSet nhằm tạo ra tập VDR từ các cặp véc tơ đặc trưng ảnh. Thuật toán EMRTrainingSet sử dụng các cặp véc tơ đặc trưng từ tập IC, tiến hành tính toán các vector khác biệt và xác định giá trị xếp hạng tương tự, từ đó xây dựng tập VDR. Thuật toán được trình bày như sau:

Thuật toán 3.2. EMR-TS (EMR Training Set)

Input:

IC: Tập hợp các cặp vector đặc trưng ảnh

Tham số và các ma trận của đồ thị EMR (C, Z, W, n, d...)

Output: Tập đầu vào, đầu ra VDR

Bước 1. Khởi tạo $VDR = \emptyset$ (VDR rỗng)

Bước 2. Với mỗi cặp vector đặc trưng $(v_1, v_2) \in IC$,

Tính vector đầu vào $x = (|v_1 - v_2|)_{1 \le i \le d} \quad _{(v_1)}$,

Tính giá trị đầu ra $\vec{r} = EMR(v_1)$,

Lấy giá trị tương ứng $r[index(v_2)]$,

Bổ sung cặp $(x, r[index(v_2)])$ vào VDR

return VDR.

Độ phức tạp thuật toán:

Giả sử n' là số lượng cặp vector trong IC và d là số chiều của mỗi vector đặc trưng:

Việc khởi tạo VDR có độ phức tạp $O(1)$.

Với mỗi cặp vector (v_1, v_2), tính toán vector đầu vào x có độ phức tạp $O(d)$; tính toán giá trị đầu ra $r = EMR(v_1)$ có độ phức tạp *O(C*n*d) + O(C^3);* lấy giá trị tương ứng $r[index(v_2)]$ có độ phức tạp $O(1)$; bổ sung cặp (x, r) vào VDR có độ phức tạp $O(1)$.

Do đó, tổng độ phức tạp của việc xử lý một cặp vector là O(d+*C*n*d+1+1) = O(C*n*d).*

Với n' cặp vector trong IC, tổng độ phức tạp là *O(n'*C*n*d)+O(C^3).* Như vậy, khi cố định EMR độ phức tạp của thuật toán phụ thuộc chủ yếu vào số lượng cặp vector trong IC.

Thuật toán EMR-TS đảm bảo rằng tất cả các cặp véc tơ trong IC được xử lý một cách hệ thống và chính xác, tạo ra một tập VDR đầy đủ và chất lượng. Tập VDR này sẽ là cơ sở dữ liệu quan trọng để tiến hành huấn luyện và đánh giá mô hình trong việc xây dựng độ đo tương tự S.

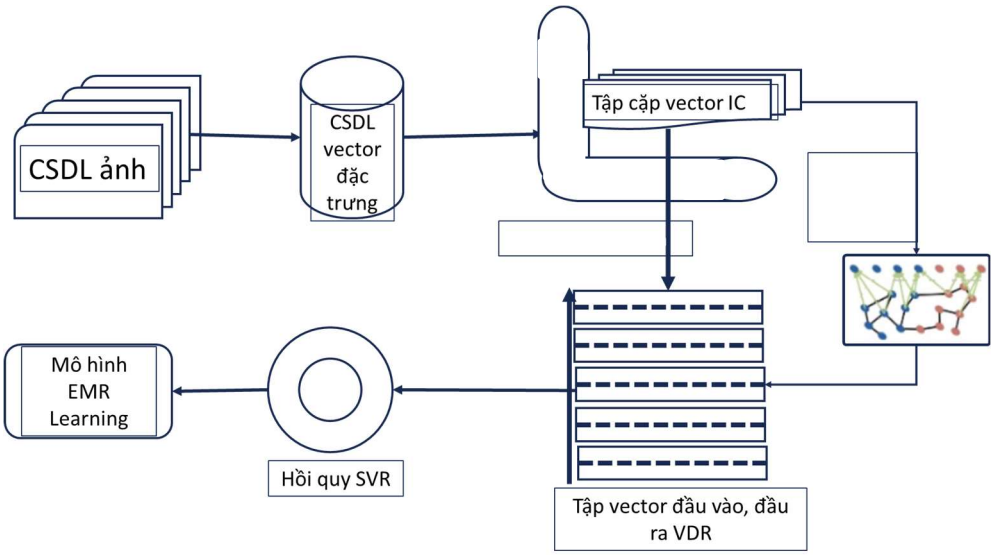
3.4 Xây dựng độ đo tương tự S của EMR learning dựa trên tiếp cận học máy hồi quy một đầu ra.

Sau khi đã xác định được tập huấn luyện VDR, luận án tiếp tục sử dụng phương pháp hồi quy để xây dựng độ đo tương tự S. Phương pháp này vận dụng thuật toán học có nhãn dưới dạng hồi quy, thay vì học phân lớp, để đảm bảo

rằng mô hình có thể dự đoán chính xác giá trị tương tự giữa các cặp ảnh. Lý do chính cho việc sử dụng học hồi quy thay vì học phân lớp là do đặc tính liên tục và số lượng lớn của các giá trị tương tự. Trong khi học phân lớp thường được sử dụng để phân loại các đối tượng vào các nhóm riêng biệt, học hồi quy lại phù hợp hơn trong việc dự đoán các giá trị liên tục, như trong trường hợp này là thứ hạng tương tự giữa các ảnh.

Một thách thức quan trọng trong quá trình này là số lượng mẫu huấn luyện có thể rất lớn vì tập hợp các cặp ảnh từ tập dữ liệu E là rất lớn. Do đó, luận án áp dụng một phương pháp học với số lượng mẫu huấn luyện giới hạn, có thể xem như là một dạng học bán giám sát. Điều này có nghĩa là mô hình được huấn luyện với một số lượng mẫu đã biết trước đầu ra, giúp giảm bớt khối lượng tính toán và tài nguyên cần thiết.

Quá trình xây dựng độ đo S thông qua học hồi quy cho phép mô hình EMR Learning có khả năng dự đoán chính xác và linh hoạt hơn trong việc xác định độ tương tự giữa các ảnh.



Hình 3.2: Các bước xây dựng dữ liệu huấn luyện cho mô hình EMR Learning

Sau đây là thuật toán xây dựng độ đo tương tự dựa trên các giá trị xếp hạng của EMR mà luận án đề xuất.

Thuật toán 3.3. EMR-LSM (Thuật toán xây dựng độ đo tương tự dựa trên các giá trị xếp hạng của EMR).

Input: $\{E_i\}_{1\le i\le n}$ là tập vector đặc trưng ảnh,

C: số lượng điểm neo (anchor points),

s: số phần tử lân cận của một vector đặc trưng,

C: số lượng điểm neo của thuật toán EMR, tham số a $\in$ (0,1) (a $\approx$ 1),

t: tham số tỉ lệ phần trăm số lượng chọn ngẫu nhiên trong tập n phần tử.

Ω_{ML}: thuật toán học máy hồi quy.

Output: độ đo tương tự ảnh Θ được xây dựng bởi thuật toán học máy.

Bước 1: Xây dựng đồ thị của EMR.

1.1.Từ tập vector đặc trưng $\{I_{,i}\}_{1\le i\le n}$.biểu diễn các ảnh trong E.

Gọi *K-means* thu được C điểm neo $\{A_c\}_{1\le c\le C}$ của tập vector $\{IC_i\}_{1\le i\le n}$

1.2. Xác định ma trận trọng số Z kích thước C x n của EMR. Z là ma trận thưa với chỉ C x s phần tử khác 0.

1.3. Từ ma trận thưa Z tính các ma trận kề W=(w_{ij}), ma trận đường chéo D = (D_{ii}) của đồ thị EMR từ các vectors $\{IC_i\}_{1\le i\le n}$. W là ma trận thưa với chỉ *n x s* số khác 0.0.

Bước 2: Xác định tập cặp vector IC.

2.1. Gọi thủ tục 3.1 (IC-S) chúng ta xác định được tập IC.

IC=IC-S

2.2. Xây dựng tập VDR.

Gọi thủ tục 3.2 EMR-TS với IC, chúng ta xác định được tập VDR. Chia VDR thành 2 tập train và test (chia ngẫu nhiên, tỷ lệ chẳng hạn là 80:20), chúng ta thu được VDR_{train}, VDR_{test}.

Bước 3: Huấn luyện mô hình học máy hồi quy trên tập VDR bởi thuật toán Ω_{ML} sẽ thu được mô hình Θ :

$\forall v_l = (v_{l,in}, v_{l,out})$, $\Theta : v_{l,in} \mapsto v_{l,out}$, ở đây $(v_{l,in}, v_{l,out}) \in VSR_{train}$.

return Θ .

Thuật toán 3.3 EMR-LSM đã mô tả chi tiết việc xây dựng độ đo tương tự S dựa trên các giá trị xếp hạng của EMR, phần tiếp theo luận án sử dụng kết quả từ thuật toán 3.3 - mô hình EMR learning để tính toán độ tương tự ảnh.

Độ phức tạp thuật toán: Độ phức tạp thuật toán của EMR-LSM sẽ tùy thuộc thuật toán hồi quy được sử dụng.

Nếu sử dụng thuật toán hồi quy SVR, do tập VDR có $K*n_b$ phần tử, nên độ phức tạp của EMR-LSM: $O(n*C+C^3)$ (độ phức tạp của EMR) + $O(n*K*n_b*d_v)$, ở đây d_v là số chiều của vector đặc trưng.

3.5 Ước lượng độ tương tự ảnh sử dụng EMR Learning

Trong thực tế, có rất nhiều bài toán yêu cầu tính toán độ tương tự giữa hai ảnh, chẳng hạn như so sánh chữ ký, so sánh khuôn mặt, hay nhận biết lá sâu bệnh. EMR Learning giúp giải quyết hiệu quả những vấn đề này bằng cách sử dụng kỹ thuật học máy để học từ kết quả xếp hạng của EMR. Qua đó, chúng ta có thể xây dựng mô hình dự đoán độ tương tự giữa các ảnh một cách chính xác và linh hoạt. Phương pháp này không chỉ khắc phục những hạn chế của EMR truyền thống mà còn mở rộng khả năng ứng dụng trong nhiều lĩnh vực thực tiễn. Sau đây, luận án mô tả quá trình sử dụng EMR Learning để tính toán độ tương tự ảnh.

Với cặp ảnh I_1, I_2 (I_1,I_2 không nhất thiết thuộc CSDL E ảnh đã cho, nhưng có ngữ nghĩa gần với các ảnh trong CSDL E).

Tính đặc trưng ảnh I_1, I_2 thu được vector v1, v2 tương ứng

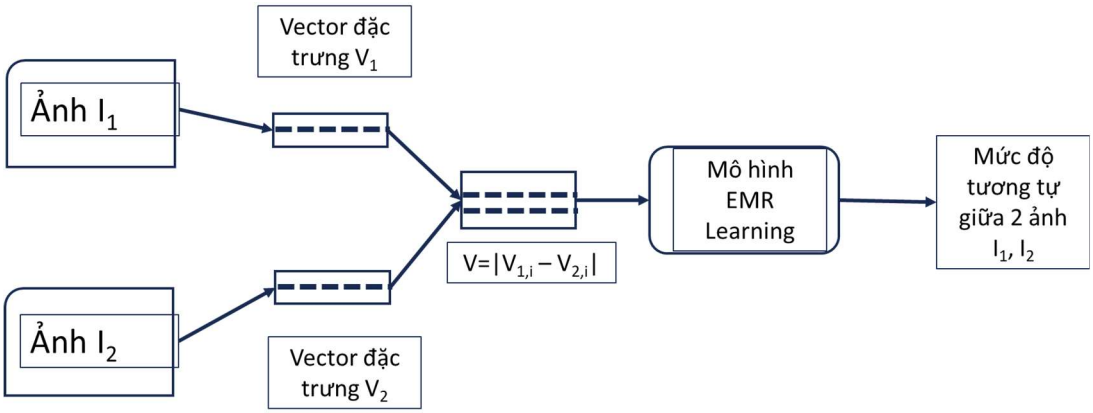
Xác định vector $\vec{v}_{SVR} = \{|v_{1,i}, v_{2,i}|\}_{1 \le i \le \dim(v_1)}$

Cho vector chạy qua mô hình EMR Learning ta thu được

$$r = EMRLearning_{(\vec{v}_{SVR})}$$

r chính là độ đo tương tự giữa cặp ảnh đầu vào I_1, I_2.

Các bước thực hiện được mô tả chi tiết tại hình 3.3



Hình 3.3: Các bước đánh giá độ tương tự của cặp ảnh bằng EMR Learning

Như vậy, khi tính độ tương tự giữa 2 ảnh I_1 và I_2 thì không cần sử dụng các phép toán ma trận và cơ sở dữ liệu có thể bổ sung thêm ảnh mới mà không cần xây dựng lại mô hình.

Sau đây là chi tiết thuật toán tính toán độ tương tự ảnh sử dụng EMR Learning.

Thuật toán 3.4. EMR-SM (Thuật toán ước lượng độ tương tự ảnh sử dụng EMR learning).

Input: 2 ảnh I_1, I_2 có thể nằm ngoài dataset ảnh E, nhưng có ngữ nghĩa cùng các ảnh của E ở giai đoạn huấn luyện.

Θ : mô hình học máy hồi quy Θ đã được huấn luyện bằng EMR-LSM.

Output: độ đo tương tự ảnh Θ được xây dựng bởi thuật toán học máy EMR-LSM

Bước 1: Tính đặc trưng kết hợp mức thấp và mức cao của I_1 và I_2.

1.1. Trích chọn đặc trưng và biểu diễn các ảnh I_1 và I_2 bằng các đặc trưng là $\{IC_i\}_{1\le i\le 2}$.

Bước 2: Dự báo độ tương tự của I_1 và I_2.

2.1. Tính vector đầu vào của Θ : $v_{in} = \left(\left|IC_1[k] - IC_2[k]\right|\right)_{1\le k\le d}$.

2.2. Dự báo độ tương tự bởi Θ và trả lại kết quả**:**

return $\Theta(v_{in})$

Tốc độ thực hiện xây dựng mô hình EMR learning phụ thuộc vào tốc độ thực hiện của kết quả dự báo (hồi quy SVR), trong đó SVR được xem là phương pháp học nhanh, hiệu quả dẫn đến việc xây dựng độ tương tự giữa cặp ảnh là phù hợp với bài toán thực tế.

Từ mô hình học độ đo tương tự EMR Learning - Thuật toán 3.3 EMR-LSM và tính toán độ tương tự ảnh sử dụng EMR Learning - Thuật toán 3.4 EMR-SM, luận án đề xuất mô hình hệ thống tra cứu ảnh dựa vào nội dung sử dụng EMR Learning theo mô tả tại hình 3.4 với các bước như sau:

Bước 1: Xây dựng độ tương tự EMR

Bước này gồm các công cụ trích rút đặc trưng ảnh theo các loại đặc trưng mức thấp, đặc trưng CNN, chuẩn hóa đặc trưng mức thấp và tìm ra bộ điểm neo. Lưu trữ dữ liệu đặc trưng và dữ liệu điểm neo để dùng cho quá trình tra cứu ảnh.

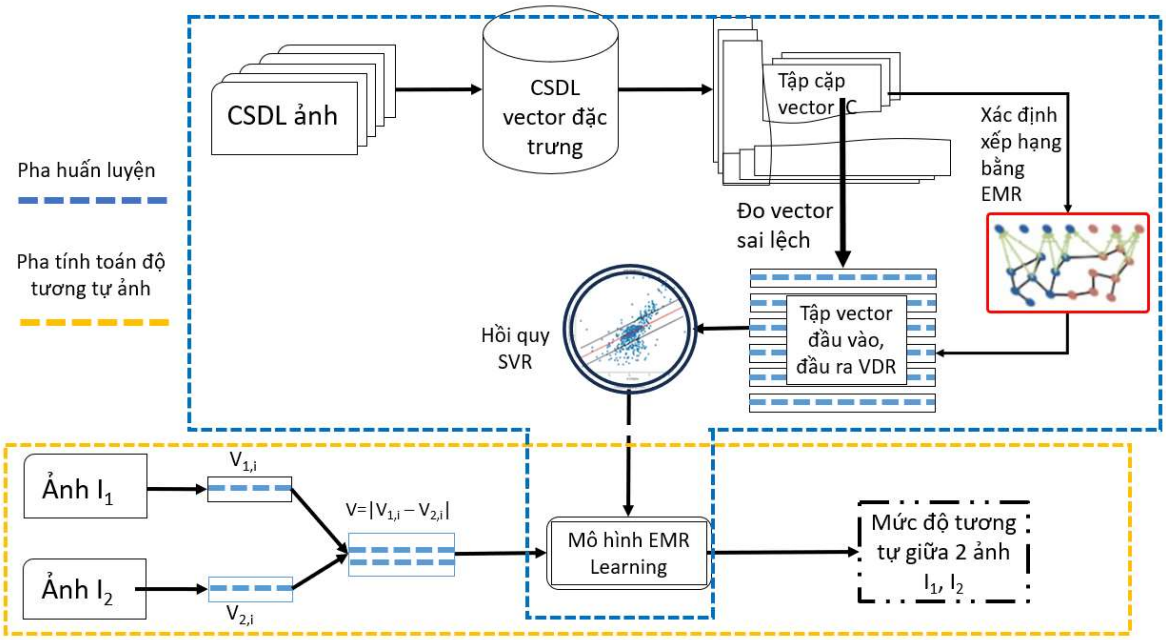
Bước 2: Học độ đo tương tự EMR Learning

Sử dụng EMR để tính toán vector xếp hạng cho từng cặp hình ảnh trong tập huấn luyện. Xây dựng tập dữ liệu gồm các cặp (vector đặc trưng, giá trị xếp hạng) để huấn luyện mô hình học máy. Sử dụng các thuật toán học máy như SVM, DNN để học mô hình mô tả mối quan hệ giữa vector đặc trưng và giá trị xếp hạng.

Mô hình được sử dụng để dự đoán giá trị xếp hạng mới cho các cặp vector chưa biết, bao gồm cả trường hợp nằm ngoài tập huấn luyện

Bước 3: Tính toán độ tương tự ảnh

Ảnh được người dùng đưa vào truy vấn và được trích rút đặc trưng với các phương pháp như ở bước 1. Sử dụng EMR Learning lựa chọn và đưa ra tập ảnh liên quan với số lượng ảnh trả về liên quan nhất do người dùng yêu cầu.



Hình 3.4: Mô hình so sánh độ tương tự ảnh sử dụng EMR Learning

Trong các phần tiếp theo, luận án đưa ra các chỉ số đánh giá hiệu quả mô hình đề xuất và tiến hành thực nghiệm tính toán độ tương tự ảnh.

3.6 Chỉ số đánh giá hiệu quả EMR Learning

Để đánh giá hiệu quả của phương pháp EMR Learning, luận án sử dụng chỉ số tương quan (R - Correlation Coefficient) [145]. Đây là chỉ số đo mức độ tương quan giữa các giá trị đầu ra thực và dự đoán của mô hình. Giá trị nằm trong khoảng từ -1 đến 1, giá trị gần 1 cho thấy mối tương quan tốt. Việc sử dụng chỉ số tương quan giúp luận án có cái nhìn tổng quan và định lượng về khả năng dự đoán của mô hình EMR Learning trong việc xác định độ tương tự giữa các ảnh.

Công thức tính toán Chỉ số tương quan Pearson (Correlation Coefficient) cho một mô hình học máy như sau:

$$R_{train} = \frac{\sum_{1\le i\le \#VDR_{train}}(r_i-\bar{r})(f_i-\bar{f})}{\sqrt{\sum_{1\le i\le \#VDR_{train}}(r_i-\bar{r})^2}\sqrt{\sum_{1\le i\le \#VDR_{train}}(f_i-\bar{f})^2}} \quad (3.3)$$

$$R_{test} = \frac{\sum_{1\le i\le \#VDR_{test}}(r_i-\bar{r})(f_i-\bar{f})}{\sqrt{\sum_{1\le i\le \#VDR_{test}}(r_i-\bar{r})^2}\sqrt{\sum_{1\le i\le \#VDR_{test}}(f_i-\bar{f})^2}} \quad (3.4)$$

Trong đó:

R_{train}: Chỉ số tương quan Pearson khi đánh giá trên bộ huấn luyện cho mô hình SVR.

R_{test}: Chỉ số tương quan Pearson khi đánh giá trên bộ huấn luyện cho mô hình SVR.

r_i: là giá trị độ tương tự cặp ảnh thứ i tính được bằng EMR.

f_i: là giá trị độ tương tự dự báo tính bằng EMR Learning.

Khi $0.7 \le R_{train} \le 1$ thì mô hình học hồi quy được xem là có khả năng học khá tốt.

Tiếp theo, luận án sẽ trình bày chi tiết các bước thực nghiệm để kiểm tra và đánh giá hiệu quả của EMR learning. Trong phần này, luận án sẽ mô tả các tập dữ liệu sử dụng, cách thức tiến hành thực nghiệm, và các kết quả thu

được. Các thí nghiệm này không chỉ giúp kiểm chứng tính khả thi của phương pháp mà còn cung cấp những thông tin quan trọng để so sánh với các phương pháp khác.

3.7. Thực nghiệm và các kết quả

3.7.1 Môi trường thực nghiệm và huấn luyện EMR Learning

Các thực nghiệm được thực hiện môi trường máy tính cá nhân.

Bảng 3.1 trình bày cấu hình của môi trường thực nghiệm trên máy tính cá nhân. Các tham số thực nghiệm được trình bày trong Bảng 3.2.

Luận án đánh giá bằng độ đo Accuracy (Độ chính xác) và Validation (kiểm thử độ chính xác của mô hình tinh chỉnh khi huấn luyện). Các tập ảnh huấn luyện được chia theo tỉ lệ 20% ảnh trong CSDL làm tập ảnh Validation, 80% ảnh trong CSDL làm tập ảnh huấn luyện.

Bảng 3.1. Môi trường thực nghiệm máy tính cá nhân

CPU	RAM	Disk
Intel® Core™ i5-6850HQ CPU @2.4 Ghz; GPU 4Ghz	32 GB DDR3	SSD 1TB

Bảng 3.2. Thời gian huấn luyện EMR Learning trên các tập dữ liệu

Thời gian huấn luyện	**Tập dữ liệu**	**Training accuracy**	**Training loss**	**Validation accuracy**
89,365 giây (>24 giờ)	Sign300HD	0.9421	0.1731	0.9215
312777 giây (>86 giờ)	Leaf30	0,9524	0,1512	0,9157
456982 giây (>126 giờ)	Corel30K	0,9163	0.1842	0,8526

3.7.2 Các tham số và kết quả thực nghiệm mô hình EMR Learning

Để đánh giá hiệu quả của mô hình EMR Learning, luận án tiến hành thực nghiệm trên các tập dữ liệu Sign300HD, Leaf30 và Corel30K. Trong đó các tham số thực nghiệm cho bước sử dụng EMR tính toán độ tương tự các véc tơ đầu vào cho mô hình EMR Learning được thể hiện tại bảng 3.3, cụ thể:

Bảng 3.3. Tham số thực nghiệm mô hình EMR Learning

TT	Tập dữ liệu	N_{best}	C	r
1	Leaf30	16	200	3
2	Sign300HD	100	2000	5
3	Corel30K	100	3000	5

Với N_{best} là số ảnh cho mỗi lần truy vấn trong 20% số lượng mẫu truy vấn lấy ngẫu nhiên, C là số điểm neo, C được chọn tùy theo bộ dữ liệu, r là số lân cận.

Để đánh giá hiệu quả của EMR Learning, tập véc tơ sai khác VDR được chia thành 2 bộ:

- Bộ huấn luyện VDR_{train}: 80% dữ liệu được sử dụng để huấn luyện mô hình EMR Learning.
- Bộ kiểm tra VDR_{test}: 20% còn lại dùng để kiểm tra độ chính xác dự đoán của mô hình.

Tiến hành huấn luyện với thuật toán học hồi quy SVR với các tham số chính [CT8]: C (Hệ số điều chỉnh)=1000, ε (Ngưỡng sai số)= 0.0001; hàm hạt nhân RBF thu được mô hình học máy EMR Learning.

Luận án tiến hành đánh giá định lượng hiệu quả của mô hình EMR Learning bằng chỉ số tương quan (R - Correlation Coefficient).

Luận án thực nghiệm huấn luyện mô hình EMR Learning trên bộ dữ liệu Sign300HD và Corel và đánh giá kết quả bằng chỉ số tương quan, kết quả thể hiện ở bảng sau:

Bảng 3.4. Kết quả đánh giá hiệu quả EMR Learning bằng chỉ số tương quan trên các tập dữ liệu

Bộ dữ liệu	Chỉ số đánh giá	
	R_{train}	R_{test}
Sign300HD	0,94	0,92
Leaf30	0.95	0.92
Corel	0.98	0.93

Điều này chứng tỏ EMR Learning có khả năng học và khái quát tốt, mang lại độ chính xác cao trong việc dự đoán độ tương tự hình ảnh. Tiếp theo, luận án tiến hành đo độ tương tự ảnh chữ ký, cụ thể với cặp ảnh chữ ký chạy qua mô hình EMR Learning cho kết quả đúng thực tế. Hình dưới đây mô tả kết quả tra cứu cặp ảnh chữ ký.

Hình 3.4: Kết quả so sánh độ tương tự ảnh chữ ký trong bộ Sign300HD

sử dụng EMR Learning

Các chỉ số đánh giá nêu trên cho phép đo độ chính xác của EMR Learning tổng quát một cách toàn diện thông qua các góc độ khác nhau như độ tương quan, độ lệch trung bình hay phương sai. Kết quả đánh giá sẽ cung cấp cơ sở tin cậy để khẳng định hiệu quả và tính khả thi của phương pháp EMR Learning.

3.8 Học xếp hạng với vấn đề nhận dạng nhãn

Trong trường hợp của các CSDL ảnh có một số giới hạn nhãn của ảnh, chúng ta đối diện với một vấn đề khác biệt so với truy vấn hình ảnh, trong đó đầu vào là hai ảnh và đầu ra là đánh giá về mức độ tương tự giữa chúng.

Đối lập với vấn đề này, là vấn đề nhận dạng nhãn của một ảnh đầu vào theo tập dữ liệu nhãn cho trước.

Tuy cả hai vấn đề liên quan đến hình ảnh, nhưng chúng có sự khác biệt trong cách tiếp cận và giải quyết. Bài toán nhận dạng thường được hiểu chính xác bởi người dùng, không có sự "mờ" như mức độ tương tự giữa hai ảnh trong bài toán tra cứu ảnh.

Tuy nhiên chúng ta có thể vận dụng một phương pháp tra cứu ảnh như EMR cho vấn đề nhận dạng nhãn của ảnh [CT5-CT7].

Giả sử chúng ta cần nhận dạng nhãn của ảnh q khi các mẫu (template) trong CSDL ảnh E có L nhãn được đánh số từ 1 đến L (để cho đơn giản trình bày, chúng ta đồng nhất ký hiệu ảnh và vector đặc trưng biểu diễn ảnh). Thuật toán nhận dạng nhãn của ảnh q theo luật vote gán nhãn của ảnh q là nhãn xuất hiện nhiều nhất trong N-top ảnh có độ tương tự cao nhất với q như sau:

Bước 1. Xác định N ảnh $\{\mathbf{E_{i1}}, \mathbf{E_{i2}}, \ldots \mathbf{E_{iN}}\}$ trong CSDL ảnh E mà có độ tương tự (theo xếp hạng độ đo khoảng cách) là lớn nhất.

Bước 2. Xác định nhãn l_0 như sau:

$$\mathbf{l_0 = vote\left(\{E_{i_k}\}_{1\le k\le N}\right) = agr\ max\{\#\{label(E_{i_k}) = l\}/1 \le l \le N\}} \quad (3.5)$$

Bước 3. Kết luận nhãn của q là l_0.

Chúng ta có thể phát biểu khái quát một chỉ số đánh giá kết quả nhận dạng nhãn ảnh của ảnh đầu vào q theo tiếp cận kiểu có khuôn mẫu (template) dựa trên một độ đo tương tự ảnh S và tập mẫu (template) ảnh E với các tham số N (lấy N-top theo giá trị tương tự lớn nhất so với q được đo bởi S) như sau và luật quyết định nhãn DS (chẳng hạn DS là luật vote, S là độ đo lấy theo độ đo khoảng cách):

$$\mathbf{Accuracy(q; S, N, DS)} = \begin{cases} \mathbf{1: l_q = DS(argmin N - top(S(q, E_i)))} \\ \mathbf{0} \end{cases} \quad (3.6)$$

$$\mathbf{Acc(S, N, DS)} = \frac{\mathbf{1}}{\mathbf{|Q|}} \sum_{\mathbf{q=1}}^{\mathbf{|Q|}} \mathbf{Accuracy(q; S, N, DS)} \quad (3.7)$$

Trong [38] thì ERR được sử dụng để đánh giá độ chính xác của mô hình AGR, ERR được tính qua chỉ số Acc như sau :

$$\mathbf{ERR(S, N, DS) = 100 - 100 \times Acc(A, N, DS)} \quad (3.8)$$

Khi đã cố định N, S và DS chúng ta sẽ chỉ viết gọn là ERR.

Để đánh giá hiệu quả của chỉ số ERR, luận án tiến hành so sánh dựa trên chỉ số ERR là tỷ lệ lỗi trung bình của các phương pháp AGR, EMR và CoEMR trên bộ dữ liệu Leaf30. Kết quả thực nghiệm ở bảng sau cho thấy CoEMR đạt độ chính xác tốt hơn, điều này làm nổi bật tính hiệu quả của việc cải thiện mô hình hóa mối quan hệ trong quá trình tra cứu ảnh.

Bảng 3.5. Bảng so sánh kết quả sử dụng chỉ số ERR

TT	Tham số	AGR	EMR	CoEMR
1.	C=1000, Nbest=100	44,26%	18.89%	**13,63%**
2.	C=1500, Nbest=100	39,43%	19.01%	**13,58%**
3.	C=2000, Nbest=100	41,55%	19.17%	**14,78%**
4.	C=2500, Nbest=100	40,49%	21.34%	**16,29%**
5.	C=3000, Nbest=100	38,33%	22.16%	**16,67%**

3.9 Kết luận chương 3

Trong chương này, luận án đã đề xuất thuật toán EMR Learning, một phương pháp dựa trên học máy và xếp hạng EMR để xây dựng các độ đo tương tự giữa hai ảnh ngay cả khi chúng nằm ngoài cơ sở dữ liệu ảnh được sử dụng để tổ chức đồ thị xếp hạng của EMR. Phương pháp này được xây dựng dựa trên kết quả xếp hạng đa tạp EMR, kế thừa những ưu điểm của EMR đã được thể hiện trong các kết quả trình bày ở Chương 2, đồng thời khắc phục hạn chế khi so sánh ảnh ngoài CSDL.

Điểm khác biệt cốt lõi của EMR Learning nằm ở việc xây dựng độ đo tương tự ảnh thông qua học bán giám sát từ giá trị xếp hạng EMR, thay vì phải thêm ảnh vào CSDL, xây dựng đồ thị quan hệ kề và tính toán véc tơ xếp hạng như cách tiếp cận truyền thống. Cụ thể, mô hình học máy được sử dụng để mô tả mối quan hệ giữa vector đặc trưng và giá trị xếp hạng, từ đó dự đoán độ tương tự cho các cặp ảnh chưa có trong CSDL.

Thuật toán EMR Learning được triển khai theo một quy trình gồm ba thành phần chính: EMR-TS (lựa chọn tập huấn luyện), EMR-LSM (học mô hình độ tương tự), và EMR-SM (dự báo độ tương tự). Cách tiếp cận này giúp loại bỏ công đoạn xây dựng đồ thị phức tạp, đơn giản hóa thuật toán và tăng khả năng mở rộng cho dữ liệu mới. Đồng thời, phương pháp này cũng nâng cao

hiệu quả trong việc ước lượng độ tương tự giữa các ảnh không thuộc tập huấn luyện ban đầu.

Khi xây dựng thuật toán học kết quả xếp hạng của EMR, EMR Learning đã vận dụng hiệu quả các thuật toán hồi quy, bao gồm SVM Regression và Random Forest Regression, trong đó tập nhãn huấn luyện được xây dựng từ cặp dữ liệu gồm vector đặc trưng ảnh và kết quả xếp hạng EMR. Trong quá trình này, CSDL vector đặc trưng có thể được cập nhật liên tục, giúp mô hình thích ứng tốt hơn với dữ liệu mới.

Phương pháp đề xuất đã được đánh giá trên các tập dữ liệu ảnh Leaf30, Sign300HD, Corel30K và VGG60K, với kết quả thực nghiệm cho thấy sự cải thiện đáng kể về độ chính xác, được đo lường qua cả đánh giá trực quan và các chỉ số khách quan. So với các phương pháp học độ đo khoảng cách khác như NCA và CMML, thuật toán EMR Learning cho thấy độ chính xác tăng đáng kể.

Bên cạnh đó, luận án cũng nghiên cứu và ứng dụng kết quả xếp hạng EMR cải tiến vào bài toán nhận dạng nhãn ảnh, theo hướng tiếp cận dựa trên khuôn mẫu (template). Kết quả thu được xác minh tính hiệu quả của phương pháp đề xuất, góp phần vào hướng nghiên cứu trong lĩnh vực tra cứu ảnh dựa trên nội dung, nhận dạng đối tượng và các ứng dụng khác trên các CSDL véc tơ trong lĩnh vực trí tuệ nhân tạo.

KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

Luận án đã trình bày khái quát một số vấn đề cơ bản liên quan đến các bài toán trong CBIR. Trên cơ sở khảo sát và phân tích các nghiên cứu liên quan về tra cứu ảnh, luận án đã tập trung nghiên cứu hai vấn đề cơ bản:

1. Kết hợp 2 xếp hạng EMR (áp dụng cho ảnh được biểu diễn bằng nhiều bộ mô tả).

2. Học độ đo tương tự EMR - EMR learning (đo độ liên quan giữa 2 ảnh có thể nằm trong hoặc ngoài CSDL ảnh).

Với vấn đề 1, luận án đạt được 2 kết quả: (1) Kết quả chính: Đề xuất thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng riêng rẽ của các bộ đặc trưng mức thấp và của bộ đặc trưng mức cao (đặc trưng nhúng, đặc trưng CNN). (2) Kết quả bổ sung: Đề xuất thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng theo từng ảnh riêng rẽ trong tra cứu ảnh với nhiều hơn 1 ảnh truy vấn.

Thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng riêng rẽ của các bộ đặc trưng mức thấp và của bộ đặc trưng mức cao (đặc trưng nhúng, đặc trưng CNN) bao gồm 3 thuật toán CoEMR-L, CoEMR-S, CoEMR-P (cho tra cứu với một ảnh truy vấn) [CT1-CT4], và thuật toán kết hợp tuyến tính và phi tuyến các kết quả xếp hạng theo từng ảnh riêng rẽ trong tra cứu ảnh với nhiều hơn 1 ảnh truy vấn EMR-FRMQ [CT8] (cho tra cứu với nhiều ảnh truy vấn) xếp hạng các véc tơ đặc trưng mức thấp, đặc trưng mức cao (véc tơ nhúng, đặc trưng CNN) sử dụng EMR, kết hợp chúng thành một xếp hạng duy nhất.

Các thuật toán kết hợp xếp hạng của EMR khi áp dụng vào CBIR đã tăng kết đáng kể độ chính xác tra cứu, và đã được minh chứng bằng thực nghiệm trên các tập dữ liệu ảnh lớn thông dụng.

Với vấn đề 2, luận án đạt được 2 kết quả: (1) Kết quả chính: Đề xuất thuật toán xây dựng độ đo tương tự ảnh từ các giá trị xếp hạng EMR theo tiếp cận

học bán giám sát. (2) Kết quả bổ sung: Đề xuất thuật toán nhận dạng ảnh theo tiếp cận nhận dạng khuôn mẫu (template) dựa trên kết quả xếp hạng dựa trên EMR cải tiến.

Về chi tiết, luận án đã đề xuất thuật toán xây dựng độ đo tương tự ảnh từ các giá trị xếp hạng EMR theo tiếp cận học bán giám sát - EMR Learning, là thuật toán đo độ tương tự giữa 2 ảnh (có thể nằm ngoài CSDL ảnh được sử dụng để huấn luyện), thuật toán được học dựa trên kết quả xếp hạng EMR và do đó kế thừa được ưu điểm của EMR [CT5-CT7].

Thuật toán EMR learning xây dựng độ đo tương tự ảnh theo tiếp cận học bán giám sát các giá trị xếp hạng EMR, vận dụng các kỹ thuật học hồi quy như SVM regression (hoặc có thể thay thế bằng Random forest regression..) để đạt được độ chính xác cao của mô hình học kết quả xếp hạng. Các kết quả thực nghiệm đã chứng tỏ thuật toán EMR Learning đã cải thiện được độ chính xác so với các thuật toán EMR gốc trên cùng bộ đặc trưng. Ngoài ra, luận án đã đề xuất thuật toán nhận dạng ảnh theo tiếp cận nhận dạng khuôn mẫu (template) dựa trên kết quả xếp hạng dựa trên EMR cải tiến. Các kết quả của luận án đã đóng góp vào lĩnh vực tra cứu ảnh dựa trên nội dung bằng cách đề xuất phương pháp kết hợp xếp hạng từ nhiều mô hình EMR và học độ đo tương tự từ giá trị xếp hạng, qua đó nâng cao độ chính xác của hệ thống CBIR.

Trong quá trình thực hiện luận án, chúng tôi (nghiên cứu sinh và tập thể giáo viên hướng dẫn) đã tổng hợp các công trình công bố quan trọng có liên quan trong phạm vi nghiên cứu của luận án, các đề xuất của luận án đã công bố ở các tạp chí/hội thảo Quốc tế/trong nước về xây dựng các vector biểu diễn ảnh cũng như các phương pháp tìm điểm neo và cải tiến thuật toán xếp hạng EMR gốc, và đã kiểm chứng hiệu quả của các thuật toán đề xuất thông qua thực nghiệm.

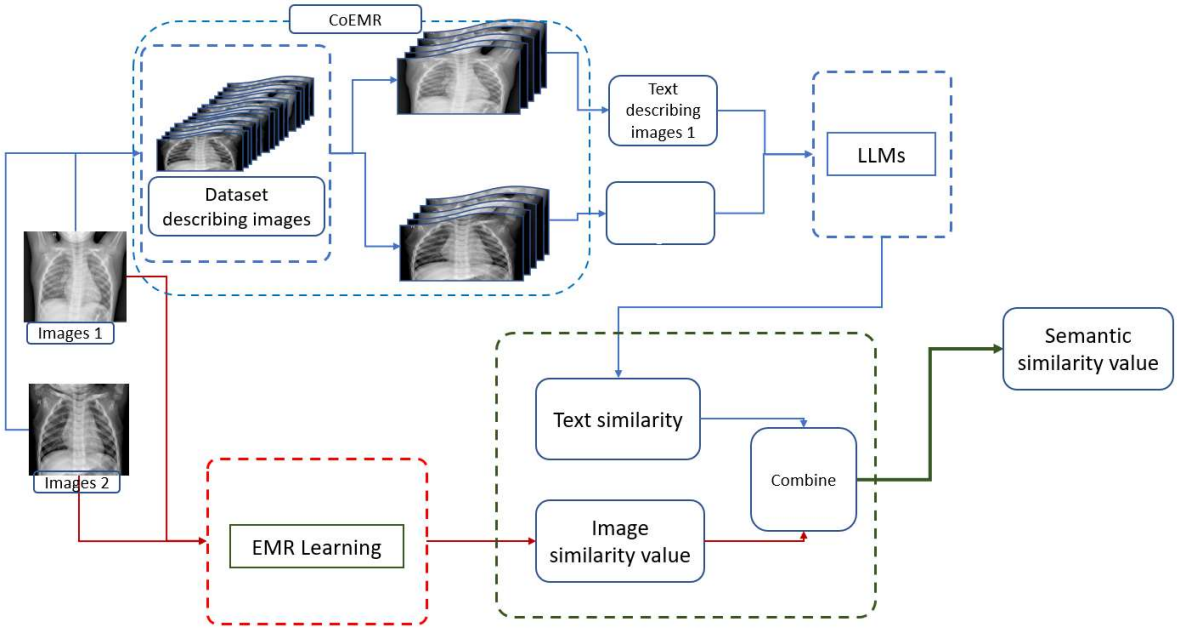
Ngoài ra, luận án sẽ tiếp tục nghiên cứu và hoàn thiện thuật toán xếp hạng đa tạp, nghiên cứu mở rộng mô hình đa truy vấn, tra cứu thông tin thị giác ở

mức ngữ nghĩa cao hơn dựa trên các thuật toán phân đoạn ảnh theo ngữ nghĩa (semantic segmentation).

Hướng nghiên cứu tiếp theo của luận án dự kiến là:

(1) Phát triển bộ dữ liệu huấn luyện gồm bộ ảnh huấn luyện cho bài toán tra cứu ảnh theo nội dung với bộ dữ liệu là các ảnh Y tế, ảnh Nông nghiệp cũng như các loại dữ liệu ảnh phục vụ cho các vấn đề Kinh tế/xã hội…

(2) Nghiên cứu áp dụng kỹ thuật EMR Learning cho các bộ dữ liệu khác nhau và kết hợp mô hình ngôn ngữ lớn (LLMs) tra cứu theo ngữ nghĩa nhằm nâng cao độ chính xác của các hệ thống tra cứu ảnh.



Hình 3.5: Hướng nghiên cứu kết hợp EMR Leaning và các mô hình ngôn ngữ lớn LLMs tính độ tương tự ngữ nghĩa ảnh.

DANH MỤC CÁC CÔNG TRÌNH KHOA HỌC CÓ LIÊN QUAN ĐẾN LUẬN ÁN

[CT1] **Huy Tran Van**, Dzung Pham Thi Kim, Huy Ngo Hoang and Quy Hoang Van, *"Enhancing the performance of manifold ranking in image retrieval using combined rank on low-level features and embedded vectors"*, Journal of Information Hiding and Multimedia Signal Processing(JIHMSP). 11(4), 172-186 (2020)

[CT2] **Trần Văn Huy**, Đào Văn Tuyết, Hoàng Văn Quý, Nguyễn Văn Đoàn, Hoàng Xuân Trung, Nguyễn Văn Quyền, Vũ Thị Khánh Toàn, Lê Đình Nghiệp, *"Nâng cao hiệu năng của đánh hạng đa tạp EMR trong truy vấn hình ảnh sử dụng phương pháp đánh hạng đa đặc trưng ảnh"*, Kỷ yếu Hội nghị Quốc gia lần thứ XV về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), 2022.

[CT3] **Trần Văn Huy**, Đào Văn Tuyết, Ngô Hoàng Huy, Phạm Thị Kim Dzung, Nguyễn Văn Quyền, Phạm Hồng Xuân, "*Kết hợp kết quả xếp hạng của nhiều EMR trong truy vấn ảnh theo nội dung*", Kỷ yếu Hội thảo khoa học Đại học Quốc tế Sài Gòn, 2023.

[CT4] **Trần Văn Huy**, Ngô Hoàng Huy, Đào Văn Tuyết, Trần Công Hùng, Phạm Thị Kim Dzung, Nguyễn Văn Quyền, Lê Đình Nghiệp, *"Xác định độ tương tự của các ảnh chữ ký dựa trên tiếp cận xếp hạng đa tạp"*, Kỷ yếu Hội nghị Quốc gia lần thứ XVI về Nghiên cứu cơ bản và ứng dụng Công nghệ thông tin (FAIR), 2023.

[CT5] **Tran Van Huy**, Dao Van Tuyet, Ngo Hoang Huy, Pham Thi Kim Dzung, Hoang Van Quy, Le Dinh Nghiep, Bui Dinh Chien, Nguyen Van Quyen *"Fusion methods of multiple EMR rankings in Content-based image retrieval"* The 8th International Conference on Applied & Engineering Physics (CAEP-8) - 2023.

[CT6] Trần Văn Huy, Ngô Hoàng Huy, Đào Văn Tuyết, Nguyễn Văn Đoàn, Hoàng Trọng Minh, Hoàng Văn Quý, Phạm Thị Kim Dzung, Nguyễn Thành Ý, Lê Đình Nghiệp *"Học không giám sát độ đo tương tự trên dữ*

liệu đa tạp của các bộ mô tả hình ảnh", Hội nghị Quốc gia về Điện tử, Truyền thông và Công nghệ thông tin (REV-ECIT), 2023.

[CT7] **Tran Van Huy**, Dao Van Tuyet, Ngo Nguyen Khoi, Pham Thi Kim Dzung, Ngo Hoang Huy, Bui Dinh Chien *"Nonlinear Fusion of multiple Efficient Manifold Rankings in Content-Based Medical Image retrieval"* International Symposium on Grids & Clouds (ISGC) TAIWAN - 2024. https://doi.org/10.22323/1.458.0002.

[CT8] **Tran Van Huy**, Ngo Hoang Huy, Dao Van Tuyet, Yasuhiro Hayashi, Ngo Nguyen Khoi, Le Ngoc Thanh *"Combining EMR rankings for multi-image queries"* International Computer Symposium (ICS) Taiwan. 2024.

TÀI LIỆU THAM KHẢO

[1] L. Y. Y. a. Q. T. Zheng, ""SIFT meets CNN: A decade survey of instance retrieval.," *IEEE transactions on pattern analysis and machine intelligence 40.5 (2017),* pp. 1224-1244, 2017.

[2] M. W. S. S. A. G. a. R. J. Arnold WM Smeulders, "Content-based image retrieval at the end of the early," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(12)*, 2000.

[3] Y. Y. B. a. G. H. LeCun, ""Deep learning."," *nature 521.7553 (2015),* pp. 436-444, 2015.

[4] W. e. a. Chen, " "Deep learning for instance retrieval: A survey."," *IEEE Transactions on Pattern Analysis and Machine Intelligence (2022).,* 2022.

[5] J. a. A. P. Janjua, " "Comparative Review of Content Based Image Retrieval Using Deep Learning."," *Intelligent Computing and Networking (2022):,* pp. 63-74, 2022.

[6] K. Z. X. R. S. a. S. J. He, "Deep Residual Learning for Image Recognition," in *in Editor: "Book Deep Residual Learning for Image Recognition" (edn.)*, 2016, pp. 770-778.

[7] F. G. T. a. O. C. Radenović, ""CNN image retrieval learns from BoW: Unsupervised fine-tuning with hard examples."," in *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands*, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part.

[8] F. G. T. a. O. C. Radenović, ""Fine-tuning CNN image retrieval with no human annotation."," *IEEE transactions on pattern analysis and machine intelligence 41.7 (2018): .,* pp. 1655-1668, 2018.

[9] D. e. a. Zhou, ""Learning with local and global consistency."," *Advances in neural information processing systems 16 (2003).,* vol. 16, pp. 595-602, 2003.

[10] X. J. Y. a. J. M. Li, " "Recent developments of content-based image retrieval (CBIR)."," *Neurocomputing 452 (2021),,* pp. 675-689, 2021.

[11] H. A. N. M. Z. a. M. A. F. Elnemr, ""Feature extraction techniques: fundamental concepts and survey."," in *Handbook of research on emerging perspectives in intelligent pattern recognition, analysis, and image processing.* , IGI Global,, 2016, pp. 264-294.

[12] U. e. a. Sharif, ""Scene analysis and search using local features and support vector machine for effective content-based image retrieval."," *Artificial Intelligence Review 52 (2019): 901-925.,* vol. 52, pp. 901-925, 2019.

[13] M. e. a. Yousuf, ""A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval."," *Mathematical Problems in Engineering 2018 (2018).,* 2018.

[14] H. e. a. Bay, ""Speeded-up robust features (SURF)."," *Computer vision and image understanding ,* vol. 110.3, pp. 346-359., 2008.

[15] S. e. a. Jabeen, ""An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model."," *PloS one 13.4 (2018): e0194526.,* 2018.

[16] A. M. J. S. K. S. A. A. D. S. J. G. Ashraf R., "Content based image retrieval by using color descriptor and discrete wavelet transform," *Journal of medical systems,* pp. 1-12, 2018.

[17] S. T. S. Pavithra L. K., "An efficient seed points selection approach in," *Cluster Computing,* pp. 1225-1240, 2019.

[18] D. N. Chaki J., "Image Color Feature Extraction Techniques: Fundamentals and Applications," in *Springer Nature*, 2020.

[19] J. S. Vinayak V., "CBIR system using color moment and color autoCorrelogram with block truncation coding," *International Journal of Computer Applications, 161(9),* pp. 1-7, 2017.

[20] R. T. A. Khokher, "A fast and effective image retrieval scheme using color, texture, and shape-based histograms," *Multimedia Tools Appl. 76 (20),* p. 21787-21809, 2017.

[21] K. S. I. Petrou M. M., Image processing: dealing with texture, John Wiley Sons, 2021.

[22] D. G. Garg M., "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Computing and Applications, 33,* pp. 1311-1328, 2021.

[23] S. T. S. Pavithra L. K., "An efficient seed points selection approach in dominant color descriptors (DCD)," *Cluster Computing, 22(4),* pp. 1225-1240, 2019.

[24] D. N. Chaki J., A beginner's guide to image shape feature extraction, CRC Press, 2019.

[25] E. M. F., "Content Based Image Retrieval for Multi-Objects Fruits Recognition using k-Means and k-Nearest Neighbor," in *International Conference on Data and Software Engineering,* , 2017.

[26] Y. e. a. Yang, ""Image retrieval via learning content-based deep quality model towards big data."," *Future Generation Computer Systems 112 (2020),* vol. 112, pp. 243-249, 2020.

[27] J. e. a. Wan, ""Deep learning for content-based image retrieval: A comprehensive study."," in *Proceedings of the 22nd ACM international conference on Multimedia.* , 2014.

[28] P. e. a. Desai, " "Hybrid approach for content-based image retrieval using vgg16 layered architecture and svm: an application of deep learning."," *SN Computer Science 2 (2021),* vol. 2, pp. 1-9, 2021.

[29] S. e. a. Gkelios, " "Deep convolutional features for image retrieval."," *Expert Systems with Applications 177 (2021): 114940.,* vol. 177, 2021.

[30] R. R. V. P. a. P. G. K. Saritha, ""Content based image retrieval using deep learning process."," *Cluster Computing 22 (2019): ,* vol. 22, pp. 4187-4200., 2019.

[31] Y. e. a. Rui, " "Relevance feedback: a power tool for interactive content-based image retrieval."," *IEEE Transactions on circuits and systems for video technology 8.5 (1998),* vol. 8.5, pp. 644-655, 1998.

[32] K. a. S. M. Rajakumar, ""MRI image retrieval using wavelet with Mahalanobis distance measurement."," *Journal of Electrical Engineering and Technology 8.5 (2013): 1188-1193.,* vol. 8.5, pp. 1188-1193, 2013.

[33] J. W. A. G. O. B. B. S. Dengyong Zhou, "Ranking on Data Manifolds," *Advances in Neural Information Processing Systems,* 2003.

[34] P. D. J. G. X. Z. a. Y. C. Xue-Qi Cheng, ""Ranking on Data Manifold with Sink Points"," *IEEE Transactions on Knowledge and Data Engineering (Volume: 25, Issue: 1),* January 2013.

[35] L. Z. H. L. X. R. M.-H. Y. Chuan Yang, "Saliency Detection via Graph-Based Manifold Ranking," in *Conference on Computer Vision and Pattern Recognition*, Portland, OR, USA, 2013.

[36] J. B. C. C. C. W. D. C. X. H. Bin Xu, "EMR: A Scalable Graph-based Ranking Model for Content-based Image Retrieval," *IEEE Transactions on Knowledge & Data Engineering,* vol. 27(1), pp. 102-114, JANUARY - 2015.

[37] K. J. B. a. V. T. Yu, " "Active learning via transductive experimental design."," in *Proceedings of the 23rd international conference on Machine learning. 2006.*, USA, 2006.

[38] M. e. a. Wang, ""Scalable semi-supervised learning by efficient anchor graph regularization."," *IEEE Transactions on Knowledge and Data Engineering 28.7 (2016): 1864-1877.,* vol. 28.7, no. IEEE, pp. 1864-1877, 2016.

[39] M. e. a. Zhao, ""A scalable sub-graph regularization for efficient content based image retrieval with long-term relevance feedback enhancement."," *Knowledge-based systems 212 (2021): 106505.,* vol. 212, 2021.

[40] S. T. Roweis, "Nonlinear Dimensionality Reduction by Locally Linear Embedding.," *Science, vol. 290, no. 5500, ,* vol. vol. 290, pp. 2323-2326, 2000.

[41] W. J. H. a. S.-F. C. Liu, " "Large graph construction for scalable semi-supervised learning."," in *Proceedings of the 27th international conference on machine learning (ICML-10). 2010.*, 2010.

[42] B. e. a. Xu, " "Efficient manifold ranking for image retrieval."," in *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval.* , 2011.

[43] H. L. F. M. Q. W. L. X. Lei Ma, ""Manifold-ranking embedded order preserving hashing for image semantic retrieval"," *Journal of Visual Communication and Image Representation, 44. doi:10.1016/j.jvcir.2017.01.014 ,* pp. 29-39, 2017.

[44] X. a. P. N. He, ""Locality preserving projections."," *Advances in neural information processing systems 16 (2003).,* vol. 16, 2003.

[45] R. Chang, "Effective Graph-Based Content-Based Image Retrieval Systems for Large-Scale and Small-Scale Image Databases," *Computer Science,* 2013.

[46] M. L. H.-J. Z. H. T. a. C. Z. J. He, "Manifold-ranking based image retrieval," in *ACM International Conference on Multimedia*, New York, NY, USA, 2004.

[47] Y. H. X. G. Y. Z. A. N. Z. JUN WU, ""Heterogeneous Manifold Ranking for Image Retrieval"," *IEEE Access,* 2017.

[48] N. S. Nira Dyn, ""Manifold-valued subdivision schemes based on geodesic inductive averaging"," *Journal of Computational and Applied Mathematics, Volume 311,* pp. Pages 54-67, 2017.

[49] Y. Y. Q. W. a. F. N. Xiaojun Chen, ""Fast Manifold Ranking With Local Bipartite Graph",," *IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 30,,* 2021.

[50] B. e. a. Wang, ""Manifold-ranking based retrieval using k-regular nearest neighbor graph."," *Pattern Recognition 45.4 (2012): .,* pp. 1569-1577, 2012.

[51] Y.-Y. T.-L. L. a. H.-T. C. Lin, " "Semantic manifold learning for image retrieval."," in *Proceedings of the 13th annual ACM international conference on Multimedia.*, 2005.

[52] H. e. a. Müller, ""Performance evaluation in content-based image retrieval: overview and proposals."," *Pattern recognition letters 22.5 (2001),* pp. 593-601, 2001.

[53] D. P. H. a. N. Sebe., "How to complete performance graphs in content-based image retrieval: add generality and normalize scope," in *IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(2)*, 2005.

[54] J. Z. W. a. R. K. Y. Chen, "Clue: Cluster-based retrieval of images by unsupervised learning.," in *IEEE Transactions on Image Processing*, 2003.

[55] M. A. F. A. K. J. a. H.-J. Z. A. Vailaya, "Image classification for content-based indexing," *Trans. Img. Proc., 10(1),* pp. 117-130, January 2001.

[56] R. F. a. W. T. F. A. Torralba, "80 million tiny images: A large data set for nonparametric object and scene recognition.," in *IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11):*, 2008.

[57] .. Cao et al, "https://paperswithcode.com/dataset/vggface2-1," 2017. [Online].

[58] Q. N. e. a. Huu, ""Graph-based semisupervised and manifold learning for image retrieval with SVM-based relevant feedback."," *Journal of Intelligent & Fuzzy Systems 37.1,* vol. 37.1, pp. 711-722, 2019.

[59] W. M. S. H. S. M. Y. Z. H. W. S. J. Jing Wang, "https://paperswithcode.com/paper/logo-2k-a-large-scale-logo-dataset-for," 2019. [Online].

[60] T. D. V. H. N. H. S. A. C. N. Q. a. Q. H. V. Trung Hoang Xuan, "“A Novel Non-Gaussian Feature Normalization Method and its Application in Content Based Image Retrieval”,," *Nonlinear Phenomena in Complex Systems journal, Volume 22 (1-17), No 1,,* pp. 1-17, 2019.

[61] S. R. Dubey, "A Decade Survey of Content Based Image Retrieval," in *IEEE Transactions on Circuits and Systems for Video Technology -arXiv:2012.00641v2 [cs.CV]*, 2021.

[62] J. A.́. J. R. a. D. L. Albert Gordo, "Deep Image Retrieval: Learning global representations for image search," *arXiv:1604.01325v2 [cs.CV] ,* 28 Jul 2016.

[63] B. a. A. D. Bhanu, "Concepts learning with fuzzy clustering and relevance feedback.," *Engineering Applications of Artificial Intelligence 15.2,* pp. 123-138, 2002.

[64] P. e. a. (Eds.), "Multi-Manifold Ranking: Using Multiple Features for Better Image Retrieval," in *PAKDD 2013, Part II, LNAI 7819, Springer-Verlag Berlin Heidelberg*, 2013, pp. 449-460.

[65] mathworks, "https://www.mathworks.com/help/matlab/math/optimizing-nonlinear-functions.html," [Online].

[66] M. a. L. Q. V. Tan, "EficientNet: Rethinking model scaling for convolutional neural networks," in *6th International Conference on Machine Learning, ICML 2019,* , 2019-June, .

[67] E. e. a. Tu, ""A novel graph-based k-means for nonlinear manifold clustering and representative selection."," *Neurocomputing 143 (2014):,* pp. 109-122, 2014.

[68] S. I. H. G. E. (. Krizhevsky A., ""Imagenet classifcation with deep convolutional neural networks,"," *Advances in Neural Information Processing Systems, ,* p. p. 1097-1105., 2012.

[69] The-Anh Pham, Van-Hao Le, Dinh-Nghiep Le, "A review of feature indexing methods for fast approximate nearest neighbor search," *5th NAFOSTED Conference on Information and Computer Science (NICS),* 2018.

[70] W. S. T. a. P. L. Zhao, ""SONG: Approximate nearest neighbor search on GPU."," *2020 IEEE 36th International Conference on Data Engineering (ICDE). IEEE, ,* no. IEEE, 2020.

[71] Keras, "Keras https://keras.io/," [Online].

[72] Tenssorflow, "https://www.tensorflow.org/," [Online].

[73] O. s. c. v. library, "OpenCV: https://www.opencv.org," [Online].

[74] T. a. M. L. Y. Matsumoto, " "Accelerating exact similarity search on cpu-gpu systems."," in *IEEE International Conference on Data Mining. IEEE, 2015.*, 2015.

[75] Facebook, "https://github.com/facebookresearch/faiss," Facebook. [Online].

[76] P. a. S. A. Arulmozhi, " "A comparative study of hash based approximate nearest neighbor learning and its application in image retrieval."," *Artificial Intelligence Review 52 (2019),* pp. 323-355., 2019.

[77] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms.," *Springer Science & Business Media.,* 2013.

[78] T. C. e. a. Havens, " "Fuzzy c-means algorithms for very large data."," *IEEE Transactions on Fuzzy Systems 20.6 (2012),* pp. 1130-1146, 2012.

[79] H. A. G. a. D. K. Verma, ""A modified intuitionistic fuzzy c-means algorithm incorporating hesitation degree."," *Pattern Recognition Letters 122 (2019),* pp. 45-52, 2019.

[80] X. J. Y. a. J. X. Wan, ""Block-based similarity search on the Web using manifold-ranking."," in *Web Information Systems-WISE, 7th International Conference on Web Information Systems Engineering*, 2006.

[81] NVIDIA, CUDA C++ Programming Guide, NVIDIA, Feb 28, 2023.

[82] Nvidia, "http://davidespataro.it/pages/Dspataro-accelleratingSCIARAfv3.html," Nvidia. [Online].

[83] J. Tölke, " "Implementation of a Lattice Boltzmann kernel using the Compute Unified Device Architecture developed by nVIDIA."," *Computing and Visualization in Science 13.1 (2010),* vol. 13.1, 2010.

[84] Y. e. a. Liu, " "Image retrieval using CNN and low-level feature fusion for crime scene investigation image database."," in *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). IEEE, 2018.*, 2018.

[85] https://scikit-learn.org/stable/auto_examples/manifold/, "https://scikit-learn.org/stable/auto_examples/manifold/," [Online].

[86] A. A. a. E. B. Vergani, ""A soft davies-bouldin separation measure."," in *2018 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE). IEEE* , 2018.

[87] X. a. Y. X. Wang, " "An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index."," in *IOP Conference Series: Materials Science and Engineering. Vol. 569.*, 2019.

[88] K. R. a. C. N. Shahapure, ""Cluster quality analysis using silhouette score."," in *IEEE 7th international conference on data science and advanced analytics (DSAA). IEEE, 2020.*, 2020.

[89] C. L. M. a. W. S. Hu, " "Fuzzy clustering validity for spatial data."," *Geo-spatial information science 11.3 (2008),* vol. 11.3, pp. 191-196, 2008.

[90] N. R. a. J. C. B. Pal, " "On cluster validity for the fuzzy c-means model.," *IEEE Transactions on Fuzzy systems 3.3 (1995): 370-379.,* vol. 3.3, pp. 370-379, 1995.

[91] S. X. B. a. Q. T. Bai, ""Scalable person re-identification on supervised smoothed manifold."," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* , 2017.

[92] X. J. Y. a. J. X. Wan, " "Towards a unified approach to document similarity search using manifold-ranking of blocks."," *Information Processing & Management 44.3 (2008): 1032-1048.,* vol. 44.3, pp. 1032-1048, 2008.

[93] L. e. a. Shen, ""A novel local manifold-ranking based K-NN for modeling the regression between bioactivity and molecular descriptors."," *Chemometrics and Intelligent Laboratory Systems 151 (2016): 71-77.,* vol. 151, pp. 71-77, 2016.

[94] M. e. a. Jian, " "Manifold ranking-based kernel propagation for saliency estimation."," in *2018 4th International Conference on Control, Automation and Robotics (ICCAR). IEEE, 2018.*, 2018.

[95] D. e. a. Tao, ""Manifold ranking-based matrix factorization for saliency detection."," *IEEE transactions on neural networks and learning systems 27.6 (2015): 1122-1134.,* vol. 27.6, no. IEEE, pp. 1122-1134, 2015.

[96] Q. J. L. a. Y. Y. Wang, ""Salient band selection for hyperspectral image classification via manifold ranking.," *IEEE transactions on neural networks and learning systems 27.6 (2016): 1279-1289.,* vol. 27.6, no. IEEE, pp. 1279-1289, 2016.

[97] R. e. a. Quan, ""Object co-segmentation via graph optimized-flexible manifold ranking."," in *Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.*, 2016.

[98] R. Y. Z. a. W. Z. He, " "Fast manifold-ranking for content-based image retrieval."," *2009 ISECS International Colloquium on Computing, Communication, Control, and Management. ,* vol. Vol.2, no. IEEE, 2009.

[99] Y. e. a. Fujiwara, ""Scaling manifold ranking based image retrieval."," in *Proceedings of the VLDB Endowment 8.4 (2014): 341-352.*, 2014.

[100] Y. H. X. G. Y. Z. a. N. Z. J. Wu, " "Heterogeneous Manifold Ranking for Image Retrieval,"," *in IEEE Access, vol. 5, pp. 16871-16884, 2017, doi: 10.1109/ACCESS.2017.2740326.,* vol. Vol.5, no. IEEE, pp. 16871-16884, 2017.

[101] J. Lee, "Introduction to topological manifolds.", Springer Science & Business Media, Vol. 202. 2010..

[102] P. Hajłasz, ""Whitney's example by way of Assouad's embedding."," in *Proceedings of the American Mathematical Society 131.11 (2003): 3463-3467.*, 2003, 131.11 (2003).

[103] J. Tenenbaum, " "Mapping a manifold of perceptual observations."," *Advances in neural information processing systems 10,* vol. 10 (1997), 1997.

[104] G. e. a. Carlsson, " "On the local behavior of spaces of natural images."," *International journal of computer vision 76 (2008): 1-12.,* vol. 76, pp. 1-12, 2008.

[105] C. S. M. a. H. N. Fefferman, ""Testing the manifold hypothesis."," *Journal of the American Mathematical Society,* vol. 29.4, pp. 983-1049., 2016.

[106] A. Q. J. &. C. G. (. Mahdi, ""DeepFeat: A bottom-up and top-down saliency model based on deep features of convolutional neural networks".," *IEEE Transactions on Cognitive and Developmental Systems, 12(1),,* vol. 12(1), no. IEEE, pp. 54-63, 2019.

[107] Y. W. H. S. S. &. B. B. Li, ""Integrating multiple deep learning models to classify disaster scene videos"."," in *In 2020 IEEE High Performance Extreme Computing Conference.*, (2020, June). .

[108] F. Q. Z. D. K. X. D. Z. Y. Z. H. .. &. H. Q. (. Zhuang, ""A comprehensive survey on transfer learning"."," in *Proceedings of the IEEE, 109(1), 43-76.*, 2020.

[109] J. M. O. V. A. &. C. V. (. Lin, "" Deep Learning-Based Descriptors for Object Instance Search"."," *Deep Learning in Object Detection and Recognition, ,* pp. 189-224, 2019.

[110] V. D. S. a. J. L. J Tenenbaum, ""A global geometric framework for nonlinear dimensionality reduction."," *science, 2000. 290(5500),* pp. p. 2319-2323, 2020.

[111] X. F. H. a. a. P. Niyogi, ""Locality preserving projections"," in *Proc. Advances in Neural Information Processing Systems, 2003: p. p 153-160.*, 2003.

[112] M. C. e. a. Cieslak, ""t-Distributed Stochastic Neighbor Embedding (t-SNE): A tool for eco-physiological transcriptomic analysis."," *Marine genomics 51 (2020): 100723.,* vol. Vol.51, 2020.

[113] W. C. M. a. K. L. Dong, ""Efficient k-nearest neighbor graph construction for generic similarity measures."," in *Proceedings of the 20th international conference on World wide web.* , 2011.

[114] T. e. a. Zhou, " "Robust visual tracking via efficient manifold ranking with low-dimensional compressive features."," *Pattern Recognition 48.8 (2015): ,* vol. 48.8, pp. 2459-2473., 2015.

[115] J. C. R. E. a. W. F. Bezdek, ""FCM: The fuzzy c-means clustering algorithm."," *Computers & geosciences 10.2-3 (1984),* Vols. 10.2-3, pp. 191-203., 1984.

[116] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters., 1973.

[117] J. e. a. He, ""Generalized manifold-ranking-based image retrieval."," *IEEE Transactions on image processing 15.10 (2006),* vol. 15.10, no. IEEE, pp. 3170-3177., 2006.

[118] K. T. S. U. a. A. I. Ahmed, " "Content based image retrieval using image features information fusion."," *Information Fusion 51 (2019): 76-99.,* vol. 51, pp. 76-99, 2019.

[119] A. e. a. Iscen, ""Efficient diffusion on region manifolds: Recovering small objects with compact cnn representations."," in *Proceedings of the IEEE conference on computer vision and pattern recognition.* , 2017.

[120] F. e. a. Yang, ""Efficient image retrieval via decoupling diffusion into online and offline processing."," in *Proceedings of the AAAI conference on artificial intelligence. Vol. 33. No. 01. 2019.*, 2019.

[121] J. M. D. a. H. J. Johnson, ""Billion-scale similarity search with gpus."," *IEEE Transactions on Big Data 7.3 (2019):,* vol. 7.3, no. IEEE, pp. 535-547., 2019.

[122] Van Quy, H, Dzung, P.T.K, Huy, N.H and Van Huy, T, "Improved EfficientNet Network for Efficient Manifold Ranking-Based Image Retrieval," in *Lecture Notes in Networks and Systems*, Ha Noi, Viet Nam, Springer, Singapore, 2023.

[123] Bac Nguyen, Francesc J. Ferri, Carlos Morell and Bernard De Baets, "An efficient method for clustered multi-metric learning," *Information Sciences,* vol. 471, pp. 149-163, 2019.

[124] Bac Nguyen, Carlos Morell and Bernard De Baets, "Supervised distance metric learning through maximization of the Jeffrey divergence," *Pattern Recognition,* vol. 64, pp. 215-225, 2017.

[125] Jacob Goldberger, Sam T. Roweis, Geoffrey E. Hinton and Ruslan Salakhutdinov, "Neighbourhood Components Analysis," in *Neural Information Processing Systems*, 2004.

[126] Bingjie Zhao and Xue-qing Li, "Course Similarity Calculation Using Efficient Manifold Ranking," in *Knowledge Science, Engineering and Management*, Springer, Cham, 2015.

[127] Bingjie Zhao and Xue-qing Li, "Course Similarity Calculation Using Efficient Manifold Ranking," in *Knowledge Science, Engineering and Management. KSEM 2015. Lecture*

Notes in Computer Science, Springer, Cham. https://doi.org/10.1007/978-3-319-25159-2_38, 2015.

[128] Quy Van Hoang, Ngo Huy Hoang, Dzung Thi Kim Pham, Tuyet Dao Van, Trong Le Van, Sergey Ablameyko and Xuan Pham Hong, "Efficient content-based image retrieval based on anchor point selection in manifold ranking with combined CNN and low-level features," *Nonlinear Phenomena in Complex Systems,* vol. Vol.26, no. No.4, pp. 366-384, 2023.

[129] Mahmut KAYA and Hasan Şakir BİLGE, "Deep Metric Learning: A Survey," *Symmetry,* 2019.

[130] Z. Ming, J. Chazalon, M. M. Luqman, M. Visani and J. -C. Burie, "Simple Triplet Loss Based on Intra/Inter-Class Metric Learning for Face Verification," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, Italy, 2017, pp. 1656-1664, doi: 10.1109/ICCVW.2017.194., 2017.

[131] J. Hu, J. Lu and Y. -P. Tan, "Discriminative Deep Metric Learning for Face Verification in the Wild," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, 2014, pp. 1875-1882, doi: 10.1109/CVPR.2014.242., 2014.

[132] Florian Schrof, Dmitry Kalenichenko and James Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015*, 2015.

[133] Dey, Sounak, Dutta, Anjan, Toledo, J. & Ghosh, Suman & Lladós and Josep & Pal, Umapada, "SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification," in *arxiv*, 2017.

[134] Jing Wei Tan and Won-Ki Jeong, "Histopathology Image Classification using Deep Manifold Contrastive Learning," p. https://arxiv.org/abs/2306.14459, 2023.

[135] e. a. Wang Meng, "Scalable semi-supervised learning by efficient anchor graph regularization.," *IEEE Transactions on Knowledge and Data Engineering.,* vol. 28.7, no. IEEE, pp. 1864-1877, 2016.

[136] F. N. a. W. Y. Wang Rong, "Fast spectral clustering with anchor graph for large hyperspectral images.," *IEEE Geoscience and Remote Sensing Letters,* Vols. 14.11 (2017): 2003-2007., 2017.

[137] J. H. a. S.-F. C. Wei Liu, "Large graph construction for scalable semi-supervised learning.," in *Proceedings of the 27th international conference on machine learning (ICML-10).*, 2010.

[138] Raju, U. S. N., Suresh Kumar, K., Haran, P., Boppana, R. S., & Kumar, N. (2020). *Content-based image retrieval using local texture features in distributed environment*. International Journal of Wavelets, Multiresolution and Information Processing, 18(01), 1941001.

[139] Oren Rippel, Manohar Paluri, Piotr Dollár, and Lubomir D. Bourdev.Metric learning with adaptive density discrimination.In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.

[140] Hoffer, E., Ailon, N. (2015). Deep Metric Learning Using Triplet Network. In: Feragen, A., Pelillo, M., Loog, M. (eds) Similarity-Based Pattern Recognition. SIMBAD 2015. Lecture Notes in Computer Science(), vol 9370. Springer, Cham. https://doi.org/10.1007/978-3-319-24261-3_7

[141] Al-Mohamade A, Bchir O, Ben Ismail MM, “Multiple Query Content-Based Image Retrieval Using Relevance Feature Weight Learning”. J Imaging. 2020 Jan 17;6(1):2. doi: 10.3390/jimaging6010002. PMID: 34460641; PMCID: PMC8321011.

[142] Simily Joseph, Kannan Balakrishnan, “Multi-Query Content Based Image Retrieval System using Local Binary Patterns”. International Journal of Computer Applications. 17, 7 (March 2011), 1-5. DOI=10.5120/2235-2857

[143] Ji Yang, Bin Xu, Binbin Lin, Xiaofei He, “Multi-Query Parallel Field Ranking for image retrieval”, Neurocomputing, Volume 135, 2014,Pages 192-202.

[144] Hsiao, Ko-Jen, Jeff Calder and Alfred O. Hero. “Pareto-Depth for Multiple-Query Image Retrieval.” IEEE Transactions on Image Processing 24 (2014): 583-594.

[145] https://libguides.library.kent.edu/SPSS/PearsonCorr

[146] http://www.cad.zju.edu.cn/home/dengcai/Data/Examples.html#EMR

[147] A. Mahdi, J. Qin and G. Crosby, "DeepFeat: A Bottom-Up and Top-Down Saliency Model Based on Deep Features of Convolutional Neural Networks," in IEEE Transactions on Cognitive and Developmental Systems, vol. 12, no. 1, pp. 54-63, March 2020, doi: 10.1109/TCDS.2019.2894561.

[148] https://blog.mlreview.com/how-to-apply-distance-metric-learning-for-street-to-shop-problem-d21247723d2a

[149] G. Salomon, A. Britto, R. H. Vareto, W. R. Schwartz and D. Menotti, "Open-set Face Recognition for Small Galleries Using Siamese Networks," *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, Niteroi, Brazil, 2020, pp. 161-166, doi: 10.1109/IWSSIP48289.2020.9145245.

Phụ lục 1

Bảng giải thích các nội dung nêu trong hình 0.1

Nội dung	Giải thích
(1.1) Đặc trưng mức thấp	- Trích rút nhanh - Mô tả thông tin địa phương - Bất biết với kích thước ảnh. - Một số đặc trưng bất biến với phép xoay ảnh, ví dụ SIFT (Scale-Invariant Feature Transform).
(1.2) Đặc trưng mức cao CNN trích xuất từ các mạng huấn luyện	- Tốc độ trích rút ảnh phụ thuộc vào độ phức tạp của mạng CNN. - Phụ thuộc vào kích thước của ảnh (ảnh thường được chuẩn hóa về kích thước cố định). - Độ phân biệt ảnh cao hơn đặc trưng mức thấp, không phụ thuộc nhiều vào chất lượng ảnh(độ nhiễu, độ sáng ảnh…). ResNet giải quyết vấn đề biến mất gradient, thường yêu cầu ảnh đầu vào có kích thước cố định (như 224x224) và vector đầu ra là một biểu diễn đặc trưng có chiều dài cố định, thường là 2048 (như ResNet-50) trước lớp phân loại cuối cùng; ảnh màu 3 kênh thường cung cấp thông tin màu sắc phong phú hơn so với ảnh xám 1 kênh, có thể dẫn đến đặc trưng mô tả tốt hơn. Mạng VGG (VGG16, VGG19): chuẩn hóa ảnh đầu vào về 224x224, và cung cấp vector đặc trưng đầu ra dày đặc với 4096 phần tử trước lớp phân loại. EfficientNet(EfficientNet-B0, EfficientNetB7…) được thiết kế để có kích thước đầu vào là 224x224, vector đặc trưng từ lớp trước lớp phân loại với 1280 chiều với EfficientNet-B0.

Nội dung	Giải thích
(1.3) Fine-tuning	Sử dụng mạng EfficentNET B0 cắt bỏ lớp cuối cùng thu được mô hình mạng và cho học qua các tập ảnh như Leaf30, Corel30K để thu được bộ vector đặc trưng CNN.
(1.4) Chuẩn hóa các thành phần vector đặc trưng	Chuẩn hóa vector đặc trưng để các thành phần vector rơi vào cùng một khoảng [a, b] giúp cân bằng đóng góp của mỗi đặc trưng vào quá trình phân loại hoặc truy vấn (điều chỉnh tỷ lệ của các thành phần để chúng nằm trong cùng một phạm vi giá trị).
(1.4.1) Chuẩn hóa min, max	Chuẩn hóa min, max là quy trình điều chỉnh các giá trị vector đặc trưng về cùng một khoảng giá trị là [0,1]. **Tính chất**: 100% số lượng của các phần tử sau khi chuẩn hóa rơi vào [0,1]. Bảo toàn thứ tự (so sánh) của các phần tử sau khi chuẩn hóa. Công thức: $$\min-\max: f_i = \{f_{i,j}\} \mapsto f_i'\{f_{i,j}'\}, f_{i,j}' = \frac{f_{i,j} - \min\limits_{E_i}\{E_{i,j}\}}{\max\limits_{E_i}\{E_{i,j}\} - \min\limits_{E_i}\{E_{i,j}\}}, \forall j = \overline{1, \dim(f_i)}$$ **Hạn chế của chuẩn hóa min-max:** - Phạm vi giá trị nhỏ: Các số trong dãy có thể rất gần nhau so với phạm vi giữa giá trị nhỏ nhất và lớn nhất. Điều này có nghĩa là phân phối của dữ liệu không có sự phân tán lớn, dẫn đến việc sau chuẩn hóa, chúng không chiếm rộng lớn trong khoảng [0, 1]. - Ngoại lệ (Outliers): Nếu có ít nhất một giá trị ngoại lệ (outlier) rất lớn hoặc rất nhỏ so với phần còn lại của dữ liệu,

Nội dung	Giải thích
	tất cả các giá trị khác, không phải là ngoại lệ, sẽ bị nén lại gần nhau sau khi chuẩn hóa. Ví dụ chuẩn hóa min-max dãy {5, 6, 7, 8, 9, 10, 11, 12, 13, 100} thu được dãy{0.00000, 0.01053, 0.02105, 0.03158, 0.04211, 0.05263, 0.06316, 0.07368, 0.08421, 1.00000}.
(1.4.2) Chuẩn hóa theo phân bố Gauss	- Chuẩn hóa Kσ (K=1,2,3...), suy từ chuẩn hóa z-score,chuẩn hóa này đặc biệt hữu ích trong các trường hợp dữ liệu tuân theo phân phối chuẩn (Gaussian). K=1: Khoảng 68% dữ liệu nằm trong phạm vi ±1. K=2: Khoảng 95% dữ liệu nằm trong phạm vi ±1. K=3: Khoảng 99.7% dữ liệu nằm trong phạm vi ±1. Do đó, khi áp dụng chuẩn hóa 3σ, "hầu hết" dữ liệu sẽ nằm giữa -1 và +1 sau khi chuẩn hóa, và chỉ có khoảng 0.3% dữ liệu là ngoại lệ nằm ngoài khoảng này. **Tính chất:** “Hầu hết” các phần tử sau khi chuẩn hóa rơi vào [-1,1] khi các thành phần tuân theo phân bố Gauss. Bảo toàn thứ tự (so sánh) của các phần tử sau khi chuẩn hóa. Ví dụ chuẩn hóa 3σ dãy {5, 6, 7, 8, 9, 10, 11, 12, 13, 100} Giả sử giá trị trung bình và độ lệch chuẩn μ=18.1, σ=28.9 (tính từ dãy 10 số trên, thay cho việc tính trên toàn bộ CSDL giá trị thành phần) thu được dãy{-0.15, -0.14, -0.1267, -0.1167, -0.1033, -0.0933, -0.08, -0.07, -0.06, 0.9433}.

Nội dung	Giải thích
(3.1.1) Độ đo tương tự	Độ đo tương nhấn mạnh việc tích hợp nhiều đặc trưng từ dữ liệu ảnh và cấu trúc đa tạp để tăng cường chất lượng của kết quả tìm kiếm ảnh dựa trên nội dung (CBIR). Phương pháp xếp hạng đa tạp được sử dụng để thúc đẩy việc xếp hạng và truy vấn, trong khi thông tin phản hồi từ người dùng giúp tinh chỉnh và cá nhân hóa kết quả tìm kiếm. Mục tiêu chung là tối ưu hóa độ chính xác và hiệu suất của hệ thống CBIR, đảm bảo kết quả tìm kiếm phản ánh chính xác và đáp ứng yêu cầu của người dùng. - Tích hợp nhiều đặc trưng ảnh và cấu trúc đa tạp để cải thiện CBIR. - Sử dụng xếp hạng đa tạp cho việc xếp hạng và truy vấn tối ưu.(MR, AGR, EMR, EMR-3sigma-OPT) - Tích hợp phản hồi người dùng để cá nhân hóa kết quả tìm kiếm. - Mục tiêu cải thiện độ chính xác và hiệu suất của hệ thống tìm kiếm. - Đảm bảo kết quả tìm kiếm phản ánh đúng nhu cầu và sở thích người dùng. - Các thuật toán về độ đo tương tự được sử dụng để xác định mức độ giống nhau giữa các đối tượng trong nhiều lĩnh vực khác nhau, bao gồm nhận dạng mẫu, truy vấn ảnh dựa trên nội dung (CBIR), hệ thống khuyến nghị và nhiều ứng dụng khác ví dụ: Khoảng cách Euclidean: Đo khoảng cách "thẳng" giữa hai điểm trong không gian nhiều chiều.

Nội dung	Giải thích
	Cosine Similarity: Đo độ tương tự giữa hai vectơ dựa trên góc giữa chúng, thường được sử dụng trong xử lý ngôn ngữ tự nhiên và truy vấn thông tin. Khoảng cách Manhattan (Taxicab or L1 norm): Đo tổng các khoảng cách theo chiều thẳng đứng và ngang giữa hai điểm. Khoảng cách Jaccard: Đo độ tương tự giữa các tập hợp dựa trên số lượng phần tử giao nhau và hợp nhau của chúng. Khoảng cách Hamming: Đo số bit khác nhau giữa hai chuỗi nhị phân, thường được sử dụng trong lĩnh vực thông tin số. Khoảng cách Mahalanobis: Tính khoảng cách giữa một điểm và một phân phối, tính đến sự tương quan giữa các biến và quy mô đơn vị của không gian đặc trưng.
(3.1.1.1) Xếp hạng đa tạp	- Xây dựng một đồ thị có trọng số cho toàn bộ tập dữ liệu, tạo ra mối liên hệ giữa các điểm dữ liệu. - Gán một giá trị xếp hạng ban đầu cho điểm truy vấn. - Các điểm liên kết lan truyền giá trị xếp hạng qua cạnh của đồ thị. - Quá trình lặp cho tới khi hội tụ. - Điểm có giá trị xếp hạng cao nhất sẽ là điểm tương tự nhất với điểm truy vấn. …. Trong đó có các thuật toán xếp hạng đã tạp: MR, AGR, EMR ,EMR-3sigma-OPT.
(3.1.1.1 d) Xếp hạng đa tạp EMR	EMR (Efficient Manifold Ranking) là một phiên bản cải tiến của thuật toán xếp hạng đa tạp MR. - Sử dụng khái niệm điểm neo thay vì các điểm dữ liệu để xây dựng đồ thị, giảm phức tạp.Tìm các điểm neo bằng phương

Nội dung	Giải thích
	pháp K-means, giảm chi phí xây dựng. Áp dụng MR trên đồ thị điểm neo thay vì toàn bộ dữ liệu.Tăng tốc quá trình lan truyền xếp hạng bằng cách làm việc trên đồ thị điểm neo. Áp dụng trong CBIR hiệu quả do giảm độ phức tạp và tính toán nhanh hơn MR. **Ưu điểm:** Giảm độ phức tạp bằng cách sử dụng điểm neo thay vì toàn bộ dữ liệu. Tính toán nhanh hơn thuật toán MR ban đầu do quy mô đồ thị nhỏ hơn.Không yêu cầu biết trước số lượng cụm của dữ liệu. **Hạn chế:** Kết quả phụ thuộc vào chất lượng của điểm neo do K-means.Chi phí tính toán ban đầu để tìm ra các điểm neo. Không xử lý được dữ liệu biến đổi, thay đổi theo thời gian.
(3.1.1.2) Độ đo Toán học (độ đo Cosine...)	- Độ đo Cosine tương tự (Cosine similarity): Đo độ tương tự giữa hai vectơ dựa trên góc giữa chúng. Đây là một độ đo thường được sử dụng trong xử lý ngôn ngữ tự nhiên và truy vấn thông tin. Công thức: Cosine similarity(A, B) = A · B / ||A|| ||B|| Trong đó A và B là hai vectơ. - Độ đo Mahalanobis tính khoảng cách giữa một điểm và một phân phối, tính đến sự tương quan giữa các biến và quy mô đơn vị của không gian đặc trưng.
(3.1.2) Độ đo khoảng cách	- Độ đo khoảng cách là thước đo mức độ khác biệt hay tương đồng giữa hai đối tượng. - Có nhiều loại độ đo khoảng cách khác nhau phụ thuộc vào bản chất dữ liệu.

Nội dung	Giải thích
	- Một số độ đo thông dụng: Euclide, Manhattan, Mahalanobis, Cosine,... - Độ đo phải đáp ứng các tiêu chuẩn của khoảng cách. - Có thể sử dụng trong nhiều lĩnh vực như học máy, xử lý ảnh,… - Là công cụ đo đánh giá độ tương tự giữa các đối tượng. - Cần lựa chọn độ đo phù hợp với bản chất dữ liệu. - Có thể kết hợp nhiều độ đo khác nhau. - Cần chuẩn hóa dữ liệu trước khi tính toán khoảng cách. - Có thể học trọng số cho mỗi thành phần để cải thiện chất lượng.
(3.1.2.1) Độ đo toán học Euclid	- Độ đo khoảng cách Euclid (Euclidean distance) đo khoảng cách "thẳng" giữa hai điểm trong không gian nhiều chiều. Đây là một độ đo căn bậc hai. **Công thức tính:** $D(x,y) = \sum_{i=1}^{k} \lvert x_i - y_i \rvert$ (k là số block) - Độ đo Euclid được sử dụng để tính toán khoảng cách giữa hai vector đặc trưng ở từng bộ đặc trưng riêng lẻ trong CBIR. - Ngoài ra, độ đo này còn được áp dụng cho nhiều bài toán khác như tìm kiếm láng giềng gần nhất, phân cụm...
(3.1.2.2) Độ đo được học	- Độ đo tự học đưa ra khả năng tự động hóa việc học các tiêu chí đánh giá khoảng cách giữa các mẫu dữ liệu, cải thiện liên tục dựa trên thông tin từ dữ liệu huấn luyện.

Nội dung	Giải thích
	- Thay vì sử dụng các độ đo cố định, hệ thống sử dụng độ đo tự học để thích ứng và tinh chỉnh mình qua quá trình phân tích dữ liệu và nhận phản hồi. - Tính linh hoạt cao của độ đo tự học cho phép nó thích ứng với đặc điểm đặc thù của dữ liệu, cải thiện đáng kể chất lượng của nhiều tác vụ phân tích dữ liệu khác nhau. - Khả năng học liên tục và tích hợp phản hồi mới giúp độ đo tự học không ngừng cải tiến, mang lại khả năng cá nhân hóa theo yêu cầu người dùng hoặc đặc trưng của tác vụ cụ thể. - Độ đo tự học giảm bớt nhu cầu can thiệp thủ công, tăng hiệu quả tính toán và hỗ trợ giải quyết các thách thức như dữ liệu không cân xứng, đóng góp vào sự phát triển của hệ thống học máy tự động và thông minh. - LMNN: tối ưu hóa khoảng cách Mahalanobis trong không gian đặc trưng để mở rộng biên giữa các lớp, giúp k-NN phân loại chính xác hơn, giảm quá khớp và cải thiện tổng quát hóa thông qua hàm mục tiêu đặc biệt, thích hợp cho dữ liệu đa dạng và phức tạp. (a) NCA: là thuật toán học có giám sát nhằm tối ưu hóa một metric khoảng cách, thường là Mahalanobis, để tăng cường hiệu suất phân loại của k-NN bằng cách điều chỉnh không gian đặc trưng sao cho các điểm cùng lớp gần nhau hơn và các điểm khác lớp xa nhau hơn, đồng thời tích hợp khả năng tự chọn lọc đặc trưng quan trọng và giảm kích thước dữ liệu một cách hiệu quả, nâng cao phân loại mềm dựa trên xác suất của hàng xóm, và áp dụng linh hoạt cho nhiều loại dữ liệu và tác vụ từ phân loại hình ảnh đến phân tích cảm xúc.

Nội dung	Giải thích
	(b) Độ đo CMML Học độ đo theo nhóm phân cụm (CMML - Clustered Multi-Metric Learning) là một phương pháp hiệu quả để giải quyết bài toán dữ liệu phân bố không đồng đều. Cơ sở ý tưởng của CMML bao gồm: - Chia dữ liệu huấn luyện thành các nhóm (cluster) không giao nhau bằng k-means. Mỗi nhóm sẽ đại diện cho một vùng của không gian tính năng. - Học một ma trận khoảng cách riêng cho mỗi nhóm dựa trên các ràng buộc triplet trong nhóm đó. - Học thêm một ma trận khoảng cách toàn cục để khiến các ma trận cục bộ khác nhau càng gần với ma trận này... Bên cạnh đó, một kỹ thuật điều chỉnh toàn cục được áp dụng nhằm bảo toàn các đặc tính chung của các cụm trong không gian metric đã được học. Triplet loss được sử dụng như một hàm mất mát để học các ma trận khoảng cách trong CMML. Cụ thể: Mỗi ràng buộc triplet bao gồm 3 điểm: điểm tâm x_i, điểm dương x_j cùng lớp với xi và điểm âm x_l khác lớp với x_i. Mục tiêu là học ma trận khoảng cách sao cho khoảng cách từ x_i đến x_j nhỏ hơn khoảng cách từ x_i đến x_l với khoảng cách biên 1. Đây chính là hàm mất mát Triplet loss cần tối thiểu hóa trong quá trình học CMML. Việc lựa chọn các ràng buộc triplet hợp lý giúp CMML học được các ma trận khoảng cách phân biệt tốt giữa các lớp. Phương pháp CMML đã được chứng minh là tăng cường tính linh hoạt và hiệu quả khi áp dụng học nhiều khoảng cách

Nội dung	Giải thích
	được học trong các ứng dụng của học máy và khai thác dữ liệu. Kỹ thuật này đã được ứng dụng thành công trong việc xử lý dữ liệu đa dạng, và đã có những ứng dụng hiệu quả trong nhiều lĩnh vực như thị giác máy tính và xử lý ngôn ngữ tự nhiên. Phương pháp CMML đã được nhóm nghiên cứu Bac Nguyen, và cộng sự đề xuất [56].

PL2: Chứng minh:

$$\frac{1}{2}\sum_{k=1}^{n}\sum_{i,j=1}^{n} W_{ij} \| \frac{1}{\sqrt{D_{ii}}} Y_{ik} - \frac{1}{\sqrt{D_{jj}}} Y_{ik} \|^2$$

$$= \sum_{k=1}^{n} \frac{1}{2}\sum_{i,j=1}^{n} W_{ij} \| \frac{1}{\sqrt{D_{ii}}} Y_{ik} - \frac{1}{\sqrt{D_{jj}}} Y_{ik} \|^2$$

$$= \sum_{k=1}^{n} \left(\frac{1}{2}\sum_{i,j=1}^{n} W_{ij} \frac{Y_{ik}^2}{D_{ii}} + \frac{1}{2}\sum_{i,j=1}^{n} W_{ij} \frac{Y_{jk}^2}{D_{jj}} - \sum_{i,j=1}^{n} W_{ij} \frac{Y_{ik}Y_{jk}}{\sqrt{D_{ii}}\sqrt{D_{jj}}} \right)$$

$$= \sum_{k=1}^{n} \left(\frac{1}{2}\sum_{i=1}^{n} Y_{ik}^2 + \frac{1}{2}\sum_{j=1}^{n} Y_{jk}^2 - \sum_{i,j=1}^{n} W_{ij} \frac{Y_{ik}Y_{jk}}{\sqrt{D_{ii}}\sqrt{D_{jj}}} \right)$$

$$= \sum_{k=1}^{n} \left(\sum_{i=1}^{n} Y_{ik}^2 - \sum_{i,j=1}^{n} W_{ij} \frac{Y_{ik}Y_{jk}}{\sqrt{D_{ii}}\sqrt{D_{jj}}} \right)$$

$$= \sum_{k=1}^{n} \left(Y_k^T Y_k - Y_k^T D^{-\frac{1}{2}} W D^{-\frac{1}{2}} Y_k \right)$$

$$= tr(Y^T Y - Y^T D^{-\frac{1}{2}} W D^{-\frac{1}{2}} Y)$$

$$= tr(Y^T (I_n - D^{-\frac{1}{2}} W D^{-\frac{1}{2}}) Y)$$

Như vậy ta có $\sum_{k=1}^{n}\sum_{i=1}^{n} \| Y_{ik} - F_{ik} \|^2 = \| Y - F \|_F^2$ chính là công thức (1.24).