

**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

TRẦN SƠN HẢI

**NGHIÊN CỨU PHƯƠNG PHÁP KẾT HỢP
CÁC ĐẶC TRƯNG MÀU SẮC HÌNH DẠNG
VÀ VỊ TRÍ ĐỂ TRUY VẤN ẢNH**

LUẬN VĂN THẠC SĨ NGÀNH TIN HỌC

**NGƯỜI HƯỚNG DẪN KHOA HỌC:
T.S. LÊ HOÀNG THÁI**

Thành phố Hồ Chí Minh - 2007

LỜI CẢM ƠN

Xin chân thành biết ơn TS. LÊ HOÀNG THÁI đã tận tình hướng dẫn, giúp đỡ tôi trong suốt quá trình làm luận văn.

Xin chân thành cảm ơn thầy cô trong khoa Công Nghệ Thông Tin đã tận tình giảng dạy và chỉ bảo tôi cùng các anh chị đồng nghiệp trong cơ quan công tác và bạn bè đã tạo điều kiện thuận lợi, đóng góp những ý kiến quý báu cho tôi trong suốt quá trình học tập và hoàn thành luận văn này.

Mặc dù rất cố gắng, song luận văn này không thể tránh khỏi những thiếu sót, kính mong được sự chỉ dẫn của các quý thầy cô và các bạn.

Tp. Hồ Chí Minh, ngày 31 tháng 12 năm 2007

Học viên Cao học Tin học khóa 14

TRẦN SƠN HẢI

LỜI MỞ ĐẦU

Trí tuệ nhân tạo (Artificial Intelligence –AI) đã phát triển hơn nửa thế kỷ qua. Có rất nhiều công trình trong và ngoài nước nghiên cứu các phương pháp sử dụng trí tuệ nhân tạo để giải các bài toán học búa trên máy tính khá thành công.

Trong phạm vi luận văn này chỉ xin trình bày nghiên cứu về phương pháp sử dụng mô hình liên mạng Noron để giải bài toán truy vấn ảnh.

Bố cục của luận văn gồm 5 chương:

Chương 1: Tổng quan về truy vấn ảnh và các hướng tiếp cận. Chương này trình bày về tình hình nghiên cứu trong và ngoài nước về lĩnh vực truy vấn ảnh cũng như các thuận lợi, khó khăn và thách thức của việc giải bài toán truy vấn ảnh.

Chương 2: Đề xuất mô hình liên mạng meta-Noron. Chương này trình bày về việc cải tiến mô hình kết hợp mạng Noron và thuật giải di truyền (NN_GA) của tác giả Lê Hoàng Thái thành mô hình liên mạng meta-Noron để giải bài toán trong trường hợp dùng bộ tham số tổng quát (m, n, L). Trong đó, m là số vector đặc trưng của mẫu X , n là số chiều của mỗi vector đặc trưng và L là số phân hoạch cho các mẫu X .

Chương 3: Một áp dụng của hệ thống liên mạng meta-Noron. Chương này trình bày quá trình xây dựng một hệ thống liên mạng cụ thể với bộ tham số ($m=4, n=5, L=3$) để giải bài toán truy vấn ảnh vùng du lịch.

Chương 4: Chương trình ứng dụng.

Chương 5: Kết luận.

MỤC LỤC

DANH MỤC BẢNG	4
DANH MỤC HÌNH VẼ VÀ BIỂU ĐỒ.....	5
Chương 1: TỔNG QUAN VỀ TRUY VẤN ẢNH VÀ CÁC HƯỚNG TIẾP CẬN.....	9
1.1. Giới thiệu chung	9
1.2. Các phương pháp truy vấn ảnh chính.....	14
1.2.1 Truy vấn theo lời chú thích (annotation, key words)	14
1.2.2 Truy vấn ảnh dựa trên nội dung (CBIR).....	14
1.2.3 Truy vấn ảnh theo đối tượng (OBIR)	17
1.2.4 Truy vấn ảnh kết hợp với máy học	17
1.3. Truy vấn ảnh dựa trên nội dung.....	18
1.3.1 Kỹ thuật truy vấn ảnh dựa trên nội dung mức 1	19
1.3.2 Truy vấn ảnh theo ngữ nghĩa.....	20
1.3.3 Các vấn đề thường gặp phải	20
1.3.4 Các cách tiếp cận hiện tại và giải pháp được chọn.....	21
1.4. Một số hệ thống truy vấn ảnh cụ thể	22
1.4.1 Truy vấn ảnh dựa trên histogram màu	22

1.4.2 Truy vấn ảnh sử dụng chỉ mục màu sắc, hình dạng và vị trí	24
1.4.3 Truy vấn ảnh dựa trên mạng Nơron.....	27
Chương 2: ĐỀ XUẤT HỆ THỐNG LIÊN MẠNG META-NƠN	34
2.1 Mạng Nơron	34
2.2 Hệ thống liên mạng meta-Nơron.....	36
2.2.1 Giới thiệu các phương pháp kết hợp.....	36
2.2.2 Kiến trúc của hệ thống liên mạng meta-Nơron.....	43
2.2.3 Quá trình huấn luyện của hệ thống liên mạng.....	48
2.2.4 Quá trình đánh giá một mẫu X qua hệ thống liên mạng	52
2.3 Hệ thống liên mạng meta-Nơron cho bài toán truy vấn ảnh.....	53
Chương 3: MỘT ÁP DỤNG CỦA HỆ THỐNG LIÊN MẠNG META-NƠN	58
3.1 Rút trích các đặc trưng của ảnh.....	58
3.2 Kiến trúc hệ thống liên mạng meta-Nơron	61
3.3 Xây dựng hệ thống liên mạng meta-Nơron.....	62
Chương 4: CHƯƠNG TRÌNH ỨNG DỤNG	70
4.1 Quá trình xử lý ảnh trước khi đưa vào hệ thống liên mạng	70
4.2 Rút trích các đặc trưng và lưu vào cơ sở dữ liệu.....	73
4.3 Tìm cấu hình mạng tối ưu.....	74

4.4 Thực hiện truy vấn tìm vùng du lịch	77
Chương 5: KẾT LUẬN.....	79
TÀI LIỆU THAM KHẢO	82
PHỤ LỤC A: Mạng Nơron Lan Truyền Ngược.....	86
PHỤ LỤC B: Cơ Sở Dữ Liệu Ảnh.....	89
PHỤ LỤC C: Cơ Sở Dữ Liệu Đặc Trưng	93

DANH MỤC BẢNG BIỂU

Bảng 1.1: Thống kê lựa chọn đặc trưng của một số hệ thống truy vấn ảnh.....	15
Bảng 1.2: Bảng thông số 25 màu cơ bản	25
Bảng 1.3: Bảng kết quả thực nghiệm hệ thống của B.G. Prasad, K.K Biswas và S.K. Gupta	26
Bảng 1.4: Kết quả hệ thống NNFIR.....	33
Bảng 3.1: Biểu diễn ảnh thành các đặc trưng	60
Bảng 3.2: Thực nghiệm phân hoạch Hạ Long theo mạng con.....	63
Bảng 3.3: Thực nghiệm phân hoạch Hà Nội theo mạng con	64
Bảng 3.4: Thực nghiệm phân hoạch Nha Trang theo mạng con.....	64
Bảng 3.5: Thực nghiệm phân hoạch qua toàn hệ thống liên mạng.....	66
Bảng 4.1: Kết quả thực nghiệm một số phương pháp.....	78
Bảng C.1: Mô tả cơ sở dữ liệu đặc trưng	94

DANH MỤC HÌNH VẼ VÀ BIỂU ĐỒ

Hình 1.1: Hệ thống truy vấn ảnh của Yahoo.....	10
Hình 1.2: Mô hình rút trích đặc trưng thời gian thực.....	11
Hình 1.3: Mô hình tách riêng cơ sở dữ liệu đặc trưng	12
Hình 1.4: Hệ thống truy vấn ảnh với cơ sở tri thức	13
Hình 1.5: Hệ thống truy vấn ảnh kết hợp màu sắc và kết cấu	22
Hình 1.6: Hai dạng lược đồ mức xám.....	23
Hình 1.7: Quá trình phân tách mức xám.....	24
Hình 1.8: Phân vùng vị trí.....	26
Hình 1.9: Hệ thống truy vấn ảnh NNFIR.....	27
Hình 1.10: Ảnh gốc	29
Hình 1.11: Ảnh qua phép biến đổi wavelet song trực giao mức 1	29
Hình 1.12: Ảnh qua phép biến đổi wavelet song trực giao mức 2.....	30
Hình 1.13: Ảnh qua phép biến đổi wavelet song trực giao mức 3.....	30
Hình 1.14: Mối quan hệ giữa khoảng cách, RBNF và UF	32
Hình 1.15: Kết quả lần truy vấn lần đầu	32
Hình 1.16: Kết quả truy vấn sau khi nhận phản hồi.....	33
Hình 2.1: Mạng Noron truyền thẳng.....	34

Hình 2.2: Mẫu X đặc trưng bởi m vector	37
Hình 2.3: Tổng thể hệ thống liên mạng	37
Hình 2.4: Quá trình xử lý của các mạng con.....	38
Hình 2.5: Phương pháp kết hợp trung bình kết quả các mạng con	39
Hình 2.6: Kết hợp mạng Noron và thuật giải di truyền (NN_GA)	40
Hình 2.7: Tổng quan mô hình kết hợp bằng hệ thống liên mạng	42
Hình 2.8: Gộp kết quả các mạng con chuẩn bị đưa vào mạng toàn cục	44
Hình 2.9: Đưa các vector R_i vào mạng kết hợp	46
Hình 2.10: Phân lớp mẫu m vector	46
Hình 2.11: Mô hình liên mạng (m, n, L).....	47
Hình 2.12: Huấn luyện mạng con RN_1 của hệ thống liên mạng meta-Noron (m, n, L)	49
Hình 2.13: Huấn luyện mạng thành phần TP_1 của hệ thống liên mạng meta-Noron (m, n, L)	51
Hình 2.14: Sơ đồ rút trích đặc trưng	54
Hình 2.15: Kiến trúc hệ thống liên mạng meta-Noron áp dụng trên ảnh	56
Hình 3.1: Sơ đồ rút trích đặc trưng	58
Hình 3.2: Ảnh đầu vào	60
Hình 3.3: Hệ thống liên mạng (m=4, n=5, L=3).....	62

Hình 3.4: Biểu đồ thực nghiệm mạng phân lớp Hạ Long.....	67
Hình 3.5: Biểu đồ thực nghiệm mạng phân lớp Hà Nội	67
Hình 3.6: Biểu đồ thực nghiệm mạng phân lớp Nha Trang.....	67
Hình 3.7: Tổng quan quá trình xử lý của hệ thống liên mạng cụ thể ($m=4$, $n=5$, $L=3$)	69
Hình 4.1: Ảnh gốc dùng truy vấn.....	70
Hình 4.2: Phần nền và phần nổi	71
Hình 4.3: Vùng 1	71
Hình 4.4: Vùng 2.....	72
Hình 4.5: Vùng 3.....	72
Hình 4.6: Vùng 4.....	73
Hình 4.7: Cập nhật đặc trưng vào cơ sở dữ liệu.....	74
Hình 4.8: Huấn luyện với số nút ẩn tối ưu	76
Hình 4.9: Kết quả truy tìm vùng du lịch	77
Hình 4.10: Chức năng thống kê xác định vùng du lịch.....	78
Hình 4.11: Biểu đồ so sánh một số phương pháp	79
Hình A.1: Mạng Noron	86
Hình A.2: Hoạt động của một Noron.....	86
Hình A.3: Một số hàm truyền	87

Hình B.1: Ảnh huấn luyện 1	89
Hình B.2: Ảnh huấn luyện 2	89
Hình B.3: Ảnh huấn luyện 3	90
Hình B.4: Ảnh huấn luyện 4	90
Hình B.5: Ảnh kiểm tra 1	91
Hình B.6: Ảnh kiểm tra 2	91
Hình B.7: Ảnh kiểm tra 3	92
Hình B.8: Ảnh kiểm tra 4	92
Hình C.1: Cơ sở dữ liệu đặc trưng	93

Chương 1: TỔNG QUAN VỀ TRUY VẤN ẢNH VÀ CÁC HƯỚNG TIẾP CẬN

1.1 Giới thiệu chung

Xử lý ảnh là lĩnh vực nghiên cứu đang phát triển không ngừng bởi tính trực quan sinh động cũng như khả năng áp dụng vào thực tế lớn. Hiện xử lý ảnh đang giành được nhiều sự quan tâm của các nhà nghiên cứu trong và ngoài nước. Trong xử lý ảnh, truy vấn ảnh có thể nói là lĩnh vực đòi hỏi sự nghiên cứu tổng hợp: nghiên cứu xử lý ảnh để rút trích các đặc trưng, áp dụng các tính toán toán học cao cấp để xác định mức độ tương đồng giữa hai ảnh và sự tổ chức sắp xếp chỉ mục cho cơ sở dữ liệu ảnh. Chính vì thế truy vấn ảnh là lĩnh vực nghiên cứu đem lại nhiều thú vị.

Hơn nữa, cùng với sự phát triển của phần mềm và phần cứng, khối lượng ảnh phát triển không ngừng và ngày càng lớn. Một số lượng lớn các ảnh đang được sử dụng ở trong thư viện ảnh số và trên web. Vì vậy nhu cầu tìm kiếm ảnh là một nhu cầu tất yếu. Hiện tại, truy vấn ảnh ứng dụng trong khá nhiều lĩnh vực như: quản lý nhãn hiệu logo, truy tìm tội phạm, ứng dụng trong y khoa, quân sự ...

Bài toán tìm kiếm ảnh trong Cơ Sở Dữ Liệu (CSDL) ảnh được đưa ra từ cuối năm 1970 và có nhiều cách giải quyết khác nhau. Cho đến ngày nay đã có rất nhiều hệ thống truy vấn ảnh (cả thương mại lẫn thực nghiệm) đã và đang được phát triển. Có hai dạng CSDL ảnh là: CSDL ảnh tĩnh và CSDL ảnh động (ảnh video...). Trong nghiên cứu này, chỉ xin xem xét đến phạm vi ảnh tĩnh.

Mục tiêu của hệ thống truy vấn ảnh là tìm ra ảnh đúng với nhu cầu của người dùng:



Hình 1.1: Hệ thống truy vấn ảnh của Yahoo

Các hệ thống truy vấn ảnh hiện nay rất đa dạng nhưng nhìn chung thì chúng được phân biệt bởi: các đặc trưng mà hệ thống rút trích từ ảnh để làm cơ sở truy vấn và phương pháp phân đoạn ảnh được sử dụng trong hệ thống truy vấn, cách thức tổ chức chỉ mục cho CSDL ảnh, độ đo sự tương đồng giữa hai ảnh. Độ đo sự tương đồng là hàm số $d: H \times H \rightarrow \mathbb{R}^+$ thỏa các tính chất sau với mọi ảnh I, J, K trong không gian các ảnh H [15]:

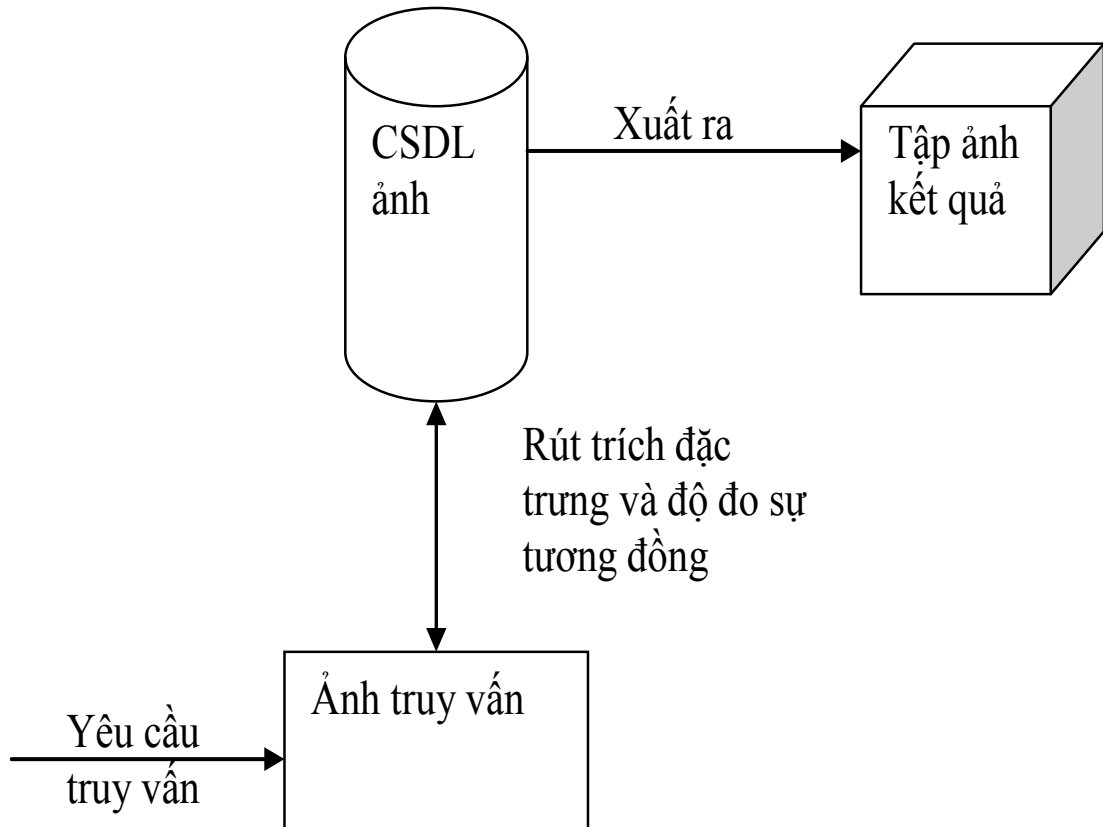
Tính tự tương đồng (self similarity): $d(I, I) = d(J, J)$

Tính đối xứng (synmetry): $d(I, J) = d(J, I)$

Tính bất đẳng (triangular inequality): $d(I, K) + d(K, J) \geq d(I, J)$

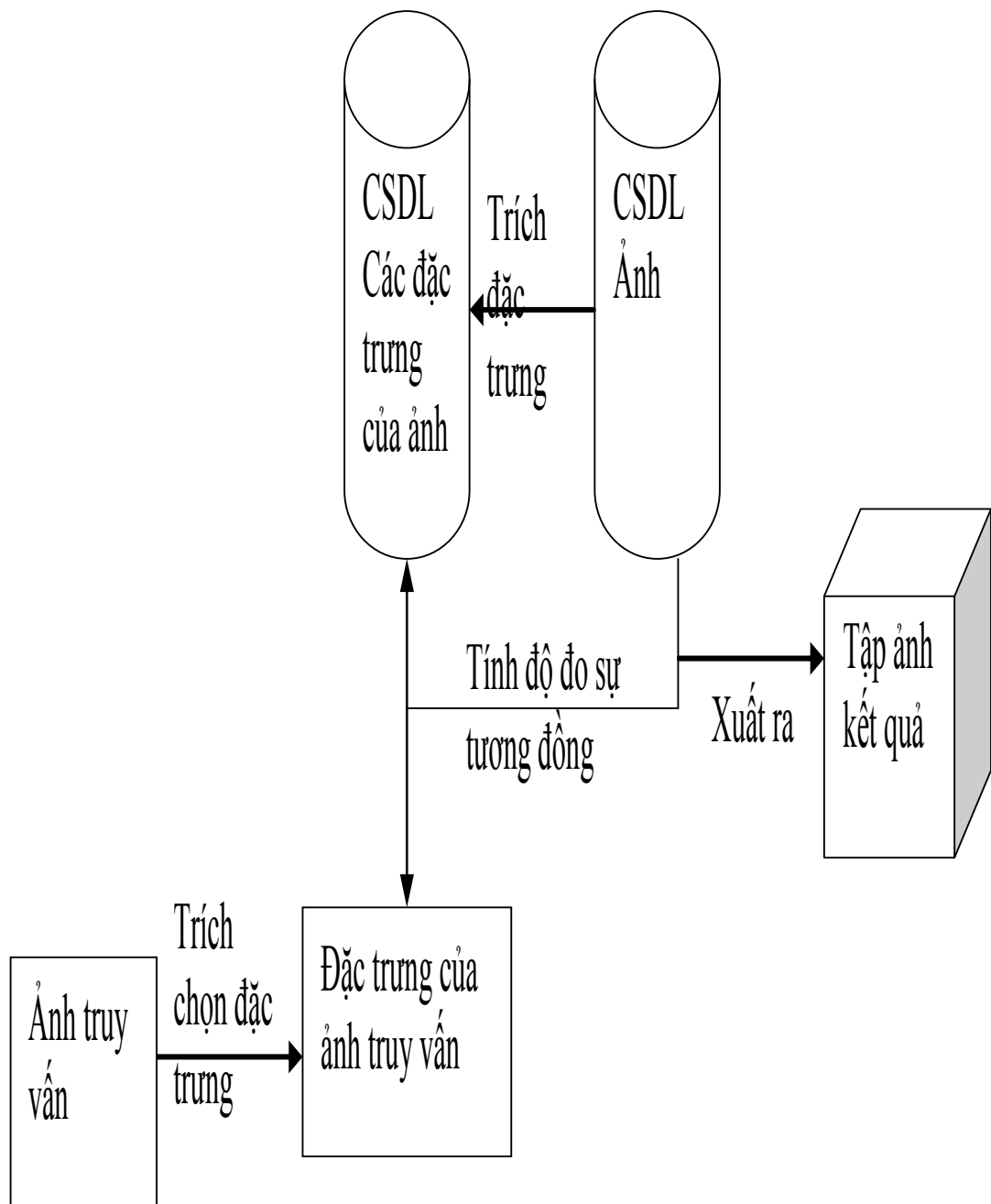
Tính tối thiểu (minimality): $d(I, J) \geq d(I, I)$

Các hệ thống truy vấn ảnh thường tuân theo một trong các mô hình sau:



Hình 1.2: Mô hình rút trích đặc trưng thời gian thực

Mô hình này đơn giản, dễ cài đặt nhưng đưa vào ứng dụng thực tế sẽ chậm vì nhược điểm: khi có một yêu cầu truy vấn ảnh hệ thống không những phải rút trích các đặc trưng của ảnh truy vấn mà còn phải rút trích các đặc trưng của ảnh trong tập ảnh của CSDL ảnh, sau đó tính toán mức độ tương đồng giữa chúng. Mô hình này thích hợp để cài đặt thử nghiệm và ghi nhận kết quả trong quá trình thực nghiệm để lựa chọn các đặc trưng cần rút trích cũng như độ đo sự tương đồng hiệu quả nhất.



Hình 1.3: Mô hình tách riêng cơ sở dữ liệu đặc trưng

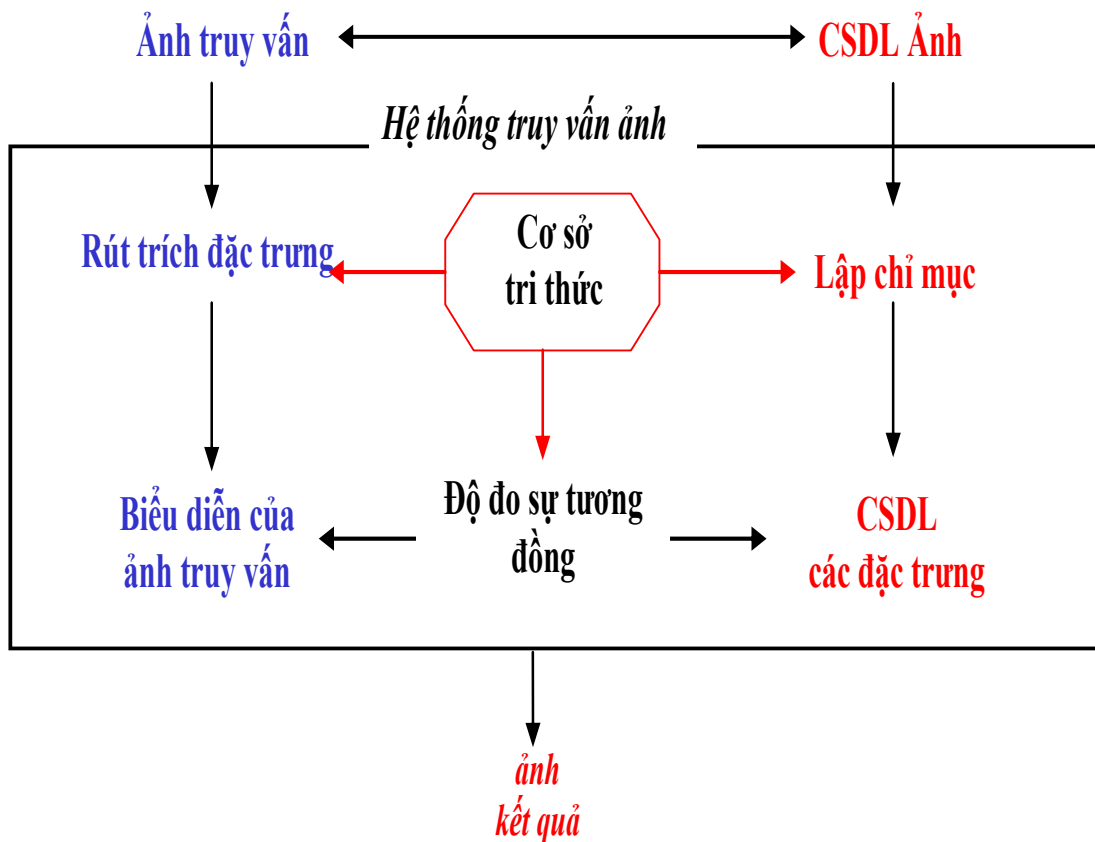
Mô hình này xây dựng nên CSDL các đặc trưng tương ứng với CSDL ảnh nên khi phát sinh một yêu cầu truy vấn, hệ thống chỉ rút trích các đặc trưng của ảnh truy vấn và so sánh với các đặc trưng trong CSDL đặc trưng nên khả năng thực thi nhanh hơn nhưng có sự phức tạp hơn trong cài đặt so với mô hình trên. Mô hình này thích hợp để xây dựng ứng dụng sau khi đã thử nghiệm và lựa chọn độ đo sự tương

đồng tối ưu, các đặc trưng cần thiết phải rút trích và phương pháp truy vấn tối ưu cho bài toán cần giải quyết. Mô hình này cần một cơ chế để đảm bảo giữa CSDL các đặc trưng và CSDL ảnh phải khớp với nhau sau một thời gian sử dụng, tránh trường hợp người dùng vô tình hay cố ý thay đổi CSDL đặc trưng không còn tương ứng với CSDL ảnh.

Tóm lại, để xây dựng hệ thống truy vấn ảnh, cần giải quyết 3 vấn đề chính sau:

- Rút trích các đặc trưng trên ảnh (Feature Exaction)
- Xác định độ đo sự tương đồng giữa hai ảnh (Similarity Measure)
- Lập chỉ mục cho CSDL ảnh (Image Indexing)

Từ đó xây dựng nên các thành phần cho hệ thống truy vấn ảnh:



Hình 1.4: Hệ thống truy vấn ảnh với cơ sở tri thức

1.2 Các phương pháp truy vấn ảnh chính

1.2.1 Truy vấn theo lời chú thích (annotation, key words)

Các từ khóa hay các lời chú thích được đưa vào để mô tả thông tin trong ảnh và các từ khóa hay lời chú thích cũng được dùng làm chỉ mục. Việc truy vấn ảnh đơn giản chỉ là sự so khớp các từ khóa đó. Cách làm này chỉ thích hợp khi các ảnh trong CSDL ảnh có nội dung không quá phức tạp. Tuy nhiên, hạn chế của cách làm này là khi CSDL ảnh lớn thì việc bổ sung từ khóa hay lời chú thích sẽ tốn nhiều chi phí tính toán và khá khó khăn. Hơn nữa, có vấn đề sẽ không thể miêu tả bằng lời chú thích mà phải thể hiện bằng thị giác của ảnh. Ngoài ra, các hệ thống truy vấn ảnh dựa trên từ khóa hay lời chú thích sẽ có sự không đồng nhất do những người sử dụng khác nhau sẽ cho ra các từ khóa khác nhau.

Bên cạnh việc truy tìm tài liệu văn bản, các bộ máy tìm kiếm của www.search.yahoo.com hay www.google.com đều hỗ trợ truy vấn ảnh dựa theo các từ khóa. Đây chính là các hệ thống truy vấn ảnh dựa trên từ khóa hay lời chú thích.

1.2.2 Truy vấn ảnh dựa trên nội dung (CBIR)

Hệ thống truy vấn ảnh dựa trên nội dung (Content-Based Image Retrieval viết tắt là **CBIR**) là hệ thống truy vấn ảnh dựa trên việc tự động rút trích một số thông tin đặc trưng trong ảnh như: màu sắc, kết cấu, vị trí, hình dạng. Phương pháp này đã được nhiều người nghiên cứu với rất nhiều cách tiếp cận khác nhau; do đó rất nhiều hệ thống truy vấn ảnh dựa trên nội dung đã ra đời như: QBIC, VisualSeek, WebSeek và BlobWorld...

Mỗi hệ thống sẽ truy vấn dựa trên một số đặc trưng nhất định nào đó và có nhiều tùy chọn khác nhau để người dùng có thể truy vấn ảnh theo màu sắc, kết cấu, hình dạng hay thậm chí là theo từ khóa. Theo thống kê tại www.aa-

lab.cs.uu.nl/cbirsurvey về mức độ lựa chọn đặc trưng để truy vấn của một số hệ thống truy vấn ảnh dựa trên nội dung cụ thể như sau:

Bảng 1.1: Thống kê lựa chọn đặc trưng của một số hệ thống truy vấn ảnh

System	Color																									Texture										Shape										Keywords	Face Detection	Layout
	Eigen image	Global histogram	Correlation histogram	Average color vector	Color coherence vector	Fixed subimage histogram	Region histogram	Dominant colors	Other	No details available	Wavelet, Fourier transform	Atomic texture features	Random fields	Local binary patterns	Edge statistics	Other	No details available	Template matching	Fourier descriptors	Elastic models	Curvature scale space	Bounding box/ellipse	Elementary descriptors	Other	No details available																							
ADL										*							*										*																					
AltaVista										*								*							*																							
Amore								*										*								*																						
BDLP								*1																		*																						
Bobworld							*				*											*																										
CANDID									*								*						*																									
C-bird		*						*2						*								*	*3	*																								
Chabot		*												*								*				*																						
CBVQ		*								*												*																										
DrawSearch				*								*						*																														
Excalibur									*					*	*		*							*																								
FIDS		*			*					*			*	*										*																								
FIR										*																																						
FOCUS					*												*					*			*																							
ImageFinder									*							*																																
ImageMiner							*			*			*						*	*		*	*	*	*	*																						
ImageRETRO		*		*				*4															*5																									
ImageRover		*			*						*		*	*										*	*	*																						
ImageScope							*	*6			*	*	*	*				*				*			*	*																						
Jacob		*									*	*	*	*		*7		*				*																										
LCPD												*	*		*7			*							*	*																						
MARS		*			*					*	*						*							*	*	*																						
MetaSEEk		*									*														*	*																						
MIR		*	*															*						*	*	*																						
NETRA						*				*								*					*		*	*																						
Photobook	*										*	*						*		*					*	*																						
Picasso						*													*		*	*	*	*	*	*																						
PicHunter		*	*		*																				*	*																						
PicToSeek		*																					*8		*	*																						
QBIC		*		*							*	*					*					*		*	*	*																						
Quicklook		*	*		*			*9		*	*	*		*		*10		*				*		*	*	*																						
SIMBA														*	*11			*		*					*	*																						
SQUID															*12					*						*																						
Surfimage	*	*								*	*		*	*	*11		*	*								*																						
SYNAPSE														*	*12											*																						
TODAI									*					*			*								*	*																						
VIR									*								*					*	*	*	*	*																						
VisualSEEk							*														*	*	*	*	*	*																						
VP																						*13				*																						
WebSEEk		*																							*	*																						
WebSeer		*						*14																	*	*																						
WISE										*															*	*																						

Hệ thống Query Based Image Content (QBIC) là hệ thống truy vấn dựa trên sự phác thảo do IBM phát triển. Người sử dụng xây dựng một phác thảo, vẽ ra và lựa chọn màu cùng kết cấu dựa theo ảnh truy vấn. Các đặc trưng màu sử dụng là giá

trị màu trung bình trong không gian RGB. Các đặc trưng về hình dạng sử dụng là dạng tròn, độ lệch tâm và hướng của trục chính. Hệ thống chỉ mục xây dựng dựa vào cấu trúc cây R^* Tree. Độ đo tương đồng về màu trung bình là $d_{avg}^2(x,y)=(x_{avg}-y_{avg})^t(x_{avg}-y_{avg})$; còn về hình dạng thì sử dụng độ đo Euclide có trọng số.

Hệ thống VisualSeek là hệ thống truy vấn dựa vào các đặc trưng trực quan của ảnh, sử dụng không gian 166 màu HSV. Sự tương đồng giữa hai ảnh được xác định theo sự tương đồng của các vùng trong ảnh. Để tiến hành truy vấn, trước tiên người dùng phải phác họa một số vùng trên ảnh. Sau đó chọn màu cho mỗi vùng, đồng thời xác định vị trí, độ lớn của vùng. Hệ thống sẽ sử dụng hàm so khớp sau để tìm các hình giống với ảnh truy vấn $d(c_q, c_t) = (c_q - c_t)^t A (c_q - c_t)$, trong đó c_q, c_t là hai tập màu của hai ảnh và $A = (a[i,j])$ là ma trận độ tương đồng của các màu.

Hệ thống WebSeek là hệ thống truy vấn ảnh trên web theo danh mục ảnh cho trước và sử dụng phép biến đổi wavelet: sưu tập ảnh sau đó phân lớp ảnh, tạo chỉ mục và cuối cùng là tìm kiếm và hiển thị. Người dùng phải chọn chủ đề trong danh mục, hệ thống sẽ sử dụng hàm so khớp sau:

$$d(h_q, h_t) = (h_q - h_t)^t A (h_q - h_t) \approx \mu_q + \mu_t - 2 \sum_{k \text{ with } h_q[k] \geq r} h_q[k] r_t[k] \quad (1.1)$$

để tìm kiếm trong chủ đề tương ứng ảnh giống nhất.

Hệ thống BlobWorld là hệ thống truy tìm theo ảnh mẫu dựa trên việc rút trích các dữ liệu điểm nguyên thủy cùng các đặc tính giống nhau về màu sắc, kết cấu và hình dạng. Màu sắc sẽ được biểu diễn dưới dạng histogram 218 màu, kết cấu đặc trưng bởi sự tương phản và tính không thẳng đứng; còn các đặc trưng hình dạng tính theo vùng với trọng tâm và hướng. Ảnh truy vấn theo sự phác thảo của các vùng riêng biệt. Hàm đo sự tương đồng là $d(h_1, h_2) = (h_1 - h_2)^T A (h_1 - h_2)$, với $A = (a_{ij})$ là ma trận đối xứng thể hiện sự tương đồng giữa màu i và j .

1.2.3 Truy vấn ảnh theo đối tượng (OBIR)

Hệ thống truy vấn ảnh dựa vào đối tượng Object Based Image Retrieval (OBIR) đang được nghiên cứu. Rõ ràng truy vấn theo đối tượng sẽ rất gần với nhu cầu thường thấy của người sử dụng và giống theo nhận thức của con người. Ví dụ tìm tất cả các vị trí đỗ xe (có biểu tượng Parking)... Tuy vậy chi phí cho việc dò tìm ra đối tượng là không nhỏ và khả năng trích ra chính xác đối tượng trong ảnh cũng không dễ bởi lẽ định nghĩa thế nào là đối tượng trong ảnh cho đúng với ý nghĩa của đối tượng trong thực tế. Vì thế chỉ riêng việc nhận diện ra đối tượng hay nhận dạng mặt người cũng là một đề tài nghiên cứu.

Đối với phương pháp này thì yếu tố quyết định là làm sao rút trích ra chính xác cùng một đối tượng trong các ảnh khác nhau (các hoàn cảnh xuất hiện khác nhau của cùng một đối tượng).

1.2.4 Truy vấn ảnh kết hợp với máy học

Để cải thiện hiệu quả truy vấn, người ta xây dựng hệ thống truy vấn ảnh kết hợp với máy học có khả năng nhận phản hồi từ người dùng. Nâng cao hiệu quả truy vấn qua quá trình huấn luyện. Ví dụ như truy vấn CSDL ảnh áp dụng mô hình Mediator Markov Model (MMM) của tác giả Nguyễn Hữu Lộc. Kết quả của hệ thống này cho thấy: trong 1048 ảnh thử nghiệm với 40 truy vấn mẫu sau khi đã có 200 phản hồi: số ảnh tìm được gần bằng 795, số ảnh tìm được đúng gần 510 và số ảnh đúng trong CSDL 670. (Nghĩa là tỷ lệ Precision=0.64 và tỷ lệ Recall là 0.76) [4].

Trong các hệ thống truy vấn ảnh áp dụng mạng Noron, dựa trên phản hồi của người dùng hệ thống thay đổi các trọng số quan hệ giữa các đặc trưng trong ảnh (như màu sắc, hình dáng...). Ví dụ như hệ thống Noron Network based Flexible Image Retrieval (NNFIR) của HYOUNG K. LEE and SUK I. YOO sử dụng hàm

Radial Basis Function (RBF) để xác lập mối quan hệ phi tuyến giữa các đặc trưng có kết quả truy vấn với tỷ lệ Precision là 81.70% và tỷ lệ recall là 54.22% [13].

Ngoài ra, các nhà nghiên cứu đã thử nghiệm nhúng Fuzzy Logic vào việc truy vấn ảnh nhưng chưa thành công vì chưa có hàm độ đo sự tương đồng thích hợp cho số mờ bởi nguyên nhân các hàm tính toán đơn giản trên số mờ thì tốt. Trong khi các tính toán cao cấp phức tạp trên số mờ vẫn còn hạn chế. Ví dụ như phương pháp nhúng Fuzzy Logic vào Truy vấn ảnh dựa trên nội dung (Embedding Fuzzy Logic in Content Based Image Retrieval) của Constantin Vertan, Nozha Boujemaa truy vấn ảnh dựa trên Fuzzy Histogram thử nghiệm với nhiều độ đo khác nhau (Zadeh, Algebraic, Lukasiewicz, Hamacher, L1...) kết quả truy vấn tỷ lệ Precision đều nhỏ hơn 70.00% [21]. Sau này, Yixin Chen và James Z. Wang (2002) đã xây dựng một hệ thống truy vấn ảnh theo vùng dựa vào Fuzzy logic sử dụng độ đo Unified feature matching (UFM) cho kết quả khá khả quan.

Bên cạnh đó, sử dụng máy học Support Vector Machine (SVM) dựa trên các phản hồi của người dùng để xác định ảnh kết quả trả ra nào là giống với ảnh truy vấn và ảnh kết quả trả ra nào là khác với ảnh truy vấn. Sau đó, hệ thống sẽ đưa ra tập kết quả chính xác hơn. Ví dụ như hệ thống “*Support Vector machine Learning for Image Retrieval*” của Lei Zhang, Fuzong, Bo Zhang cho tỷ lệ recall là 0.743. Phương pháp này đòi hỏi cơ chế ngăn chặn sự phản hồi sai của người dùng, đồng thời cần có một quá trình huấn luyện cho máy học thì hệ thống mới hoạt động tốt được.

1.3 Truy vấn ảnh dựa trên nội dung

Truy vấn ảnh dựa trên nội dung là sự truy vấn ảnh trong CSDL ảnh bằng cách tự động rút trích các đặc trưng cơ bản hay ngữ nghĩa trong ảnh truy vấn và so sánh với các đặc trưng của các ảnh trong CSDL. Để rút ra các đặc trưng trong ảnh người ta phải sử dụng các phép biến đổi như Wavelet, Fourier, DCT, hay các bộ lọc Gabor, bộ lọc trung bình,...

1.3.1 Kỹ thuật truy vấn ảnh dựa trên nội dung mức 1

Truy vấn ảnh dựa trên nội dung ở mức 1 là sự truy vấn dựa trên các đặc trưng cơ bản như màu sắc, hình dạng, kết cấu, vị trí của các thành phần trong ảnh. Ví dụ: “tìm các hình có chứa ngôi sao vàng năm cánh” hay “tìm các hình có số điểm ảnh màu đỏ chiếm $\frac{3}{4}$, còn lại là màu xanh”... Truy vấn ảnh ở mức độ này được áp dụng rộng rãi trong thực tế: tìm bản vẽ thiết kế, ứng dụng quản lý đăng ký logo (người dùng đưa vào một ảnh muốn đăng ký làm logo cho tổ chức của mình, hệ thống sẽ kiểm tra xem ảnh logo đó đã có người đăng ký hay chưa nếu đã được đăng ký hệ thống sẽ trả ra ảnh logo giống với ảnh muốn đăng ký trong cơ sở dữ liệu các ảnh logo đã được đăng ký của hệ thống)...

Màu sắc là một yếu tố rất quan trọng trong ảnh, có rất nhiều phương pháp truy vấn ảnh dựa trên sự tương tự về màu sắc. Thường các phương pháp truy vấn ảnh dựa theo màu sắc dùng histogram màu. Ví dụ hệ thống truy vấn ảnh dựa trên histogram của Lâm Thị Ngọc Châu, Trường Đại Học Khoa Học Tự Nhiên, Khoa Công Nghệ Thông Tin, 2003 [1]. Các hệ thống này thường ít bỏ sót, nghĩa là ảnh cần tìm luôn xuất hiện trong tập ảnh kết quả trả ra nhưng tỷ lệ bất nhảm rất cao hơn 50%.

Về kết cấu, việc truy vấn ảnh dựa trên kết cấu có vẻ là không hiệu quả nhưng có thể dùng nó để phân biệt các vùng hay đối tượng có màu sắc tương đồng như bầu trời và biển cả thì rất tốt...

Hình dạng là khái niệm có định nghĩa tốt hơn kết cấu và các đối tượng tự nhiên có thể phân biệt nhau bởi hình dạng. Các đặc trưng về hình dạng được sử dụng thường phải độc lập về kích thước và hướng. Ví dụ đặc trưng về tỷ số giữa chu vi và diện tích, đường biên, hình dạng tròn...

1.3.2 Truy vấn ảnh theo ngữ nghĩa

Truy vấn ảnh dựa theo ngữ nghĩa có hai cấp độ: ngữ nghĩa cấp thấp dựa trên việc rút trích một số đặc tính logic như “tìm tất cả các hình có tháp Eiffel”; còn ngữ nghĩa cấp cao thì dựa trên các thuộc tính trừu tượng như “tìm các hình có người châu Á đang khiêu vũ”.

Phần lớn các kỹ thuật truy vấn ảnh trong CBIR hiện nay đều dựa trên các đặc trưng cơ bản ở mức 1. Đã có một số công trình truy vấn ảnh dựa theo ngữ nghĩa cấp thấp như phương pháp của Ranta và Grimson dùng các mối liên hệ về màu sắc rút ra từ sự phân giải cấp thấp của ảnh để xây dựng nên các mẫu do người dùng định nghĩa.

Còn các công trình nghiên cứu truy vấn ảnh dựa trên các đặc trưng ngữ nghĩa cấp cao rất ít. Hiện tại chỉ có một số công trình đang nghiên cứu để nhận biết ảnh hoàng hôn và bình minh hay là ảnh đó thể hiện thời tiết lạnh hay ấm áp...

1.3.3 Các vấn đề thường gặp phải

Hệ thống truy vấn ảnh dựa vào ngữ nghĩa thường gặp phải vấn đề về rút trích và nhận dạng đặc trưng ngữ nghĩa trong ảnh và độ tương đồng tổng thể giữa các đặc trưng ngữ nghĩa với nhau. Ví dụ ảnh 1 có bóng đèn tròn màu đỏ, ảnh 2 có bóng đèn tròn màu cam, ảnh ba có bóng đèn elíp màu đỏ. Với trường hợp này việc rút trích chính xác các ngữ nghĩa bóng đèn hình gì và màu gì trong mỗi ảnh là không dễ. Hơn nữa, ảnh 1 sẽ giống ảnh 2 hay ảnh 1 giống ảnh 3 cũng là một vấn đề hệ thống cần xử lý. Vì ảnh 1 khác về màu và giống nhau về hình dạng đèn với ảnh 2 nhưng ảnh 1 lại khác về hình dạng đèn và giống về màu với ảnh 3.

Phương pháp truy vấn ảnh dựa trên các đặc trưng cơ bản đã chứng minh được phần nào hiệu quả của nó qua các hệ thống CBIR mức 1 đang được triển khai QBIC, VisualSeek, BlobWorld,... với hiệu quả truy vấn tương đối. Nhưng mỗi một

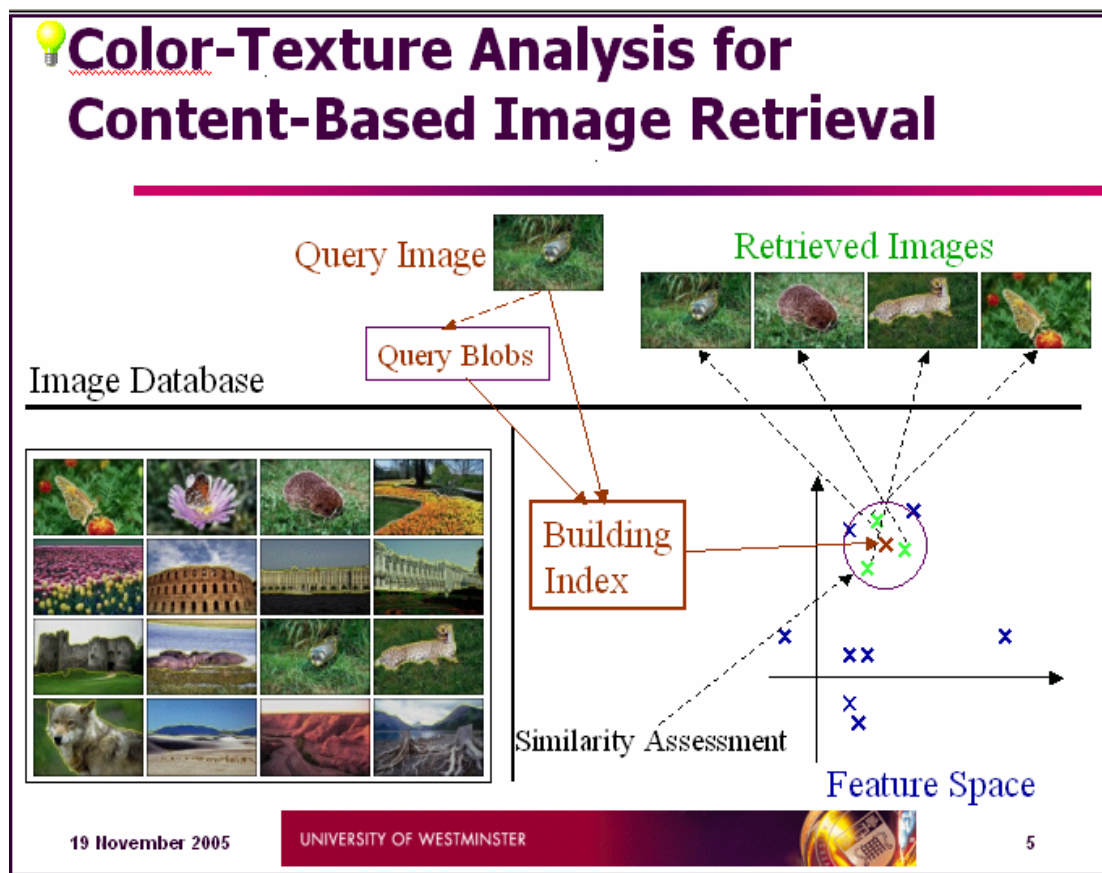
cách lựa chọn đặc trưng cơ bản nào dùng để truy vấn thường cũng chỉ tối ưu trong một số loại ảnh này mà không tốt trong loại ảnh khác. Cụ thể như, nếu dùng histogram sẽ có vô số ảnh khác nhau nhưng lại có cùng histogram, ngược lại nếu dùng hình dạng thì sẽ hạn chế khi thực hiện truy vấn trong một loại tập ảnh có sự khác biệt ít về hình dạng (như tập ảnh về các loại cá).

Ngoài ra, các hệ thống đánh chỉ mục cho dữ liệu văn bản thông thường không thể áp dụng cho việc xây dựng chỉ mục cho CSDL ảnh. Áp dụng hệ thống chỉ mục cho dữ liệu nhiều chiều R^* -tree, TV-tree và SS^+ -tree vào đánh chỉ mục cho CSDL ảnh cũng có những hạn chế nhất định về chi phí tính toán. Vì vậy để triển khai một hệ thống truy vấn ảnh hoàn thiện trong thực tế cần xây dựng hệ thống chỉ mục nhiều thích hợp.

1.3.4 Các cách tiếp cận hiện tại và giải pháp được chọn

Cùng với sự phát triển mạnh mẽ phép biến đổi wavelet, cách tiếp cận truy vấn ảnh dựa trên phép biến đổi này cũng phát triển theo. Một số công trình nghiên cứu theo hướng này đã ra đời và vẫn đang tiếp tục phát triển để cho ra những kết quả truy vấn chính xác hơn. Song, sử dụng phép biến đổi wavelet, cũng như Fourier hay DCT trước đây đòi hỏi chi phí tính toán cao. Cho nên để giải quyết vấn đề này, người ta thường chọn phép biến đổi wavelet haar để giảm bớt chi phí tính toán trong các hệ thống truy vấn ảnh. Ví dụ hệ thống áp dụng biến đổi wavelet vào truy vấn ảnh trong luận văn của tác giả Nguyễn Văn Điều thời gian tìm kiếm 10 ảnh trong 1000 ảnh dùng độ đo $l^1(Z)$ trên V_4 mất 3 phút với kết quả 85% [2].

Một hướng tiếp cận khác khá mới mẻ và đang được quan tâm là kết hợp các đặc trưng cơ bản để truy vấn ví dụ như phương pháp kết hợp màu sắc và kết cấu của tác giả Anh-Minh Hoàng. Phương pháp này đang trong giai đoạn hoàn thiện:



Hình 1.5: Hệ thống truy vấn ảnh kết hợp màu sắc và kết cấu

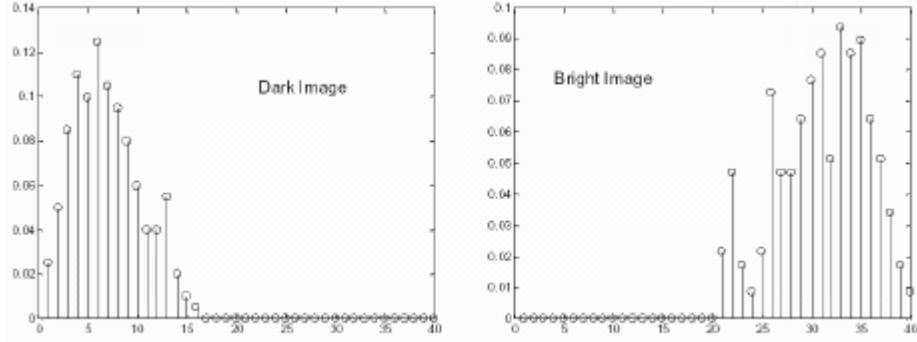
Nghiên cứu trên cho thấy việc kết hợp hai hay nhiều đặc trưng để truy vấn là phương pháp hoàn toàn khả thi. Trong phạm vi nghiên cứu của luận văn này sẽ kết hợp các đặc trưng về màu sắc, hình dạng và vị trí để truy vấn ảnh. Tuy việc kết hợp các đặc trưng sẽ cải thiện đáng kể kết quả truy vấn nhưng sẽ gặp khó khăn trong việc tìm và chọn độ đo sự tương đồng giữa 2 ảnh theo nhiều đặc trưng và xây dựng hệ thống chỉ mục cho CSDL ảnh truy vấn dựa trên nhiều đặc trưng.

1.4 Một số hệ thống truy vấn ảnh cụ thể

1.4.1 Truy vấn ảnh dựa trên histogram màu [1]

Histogram của một ảnh là một hàm cung cấp tần suất xuất hiện của mỗi mức xám. Ta có thể biểu diễn histogram trong hệ trục Oxy với trục hoành Ox biểu diễn

số mức xám từ 0 đến L (số mức xám), trục tung Oy biểu diễn số điểm ảnh cho một mức xám tương ứng.



Hình 1.6: Hai dạng lược đồ mức xám

Nghĩa là, histogram của ảnh số có mức xám $[0, L]$ là một hàm rời rạc $h(r_k) = n_k$ với r_k : mức xám thứ k và n_k là số pixel có mức k .

Histogram thường sẽ được chuẩn hoá thành $p(r_k) = n_k/n$ với n là tổng số pixel trong ảnh. Và lúc này ta có : $\left(\sum_{k=0}^L p(r_k) = 1 \right)$ (1.2)

Độ đo giữa 2 histogram gọi là độ đo phần giao cặp histogram (histogram intersection measure) của ảnh q và ảnh t là:

$$v_{q,t} = 1 - \frac{\sum_{m=0}^L \min(p^q(m), p^t(m))}{\min(\sum_{m=0}^L p^q(m), \sum_{m=0}^L p^t(m))} \quad (1.3)$$

Hệ thống truy vấn ảnh của tác giả Lâm Thị Ngọc Châu sẽ xám hoá ảnh màu RGB về ảnh xám 256 mức từ 0..255. Theo công thức mức xám:

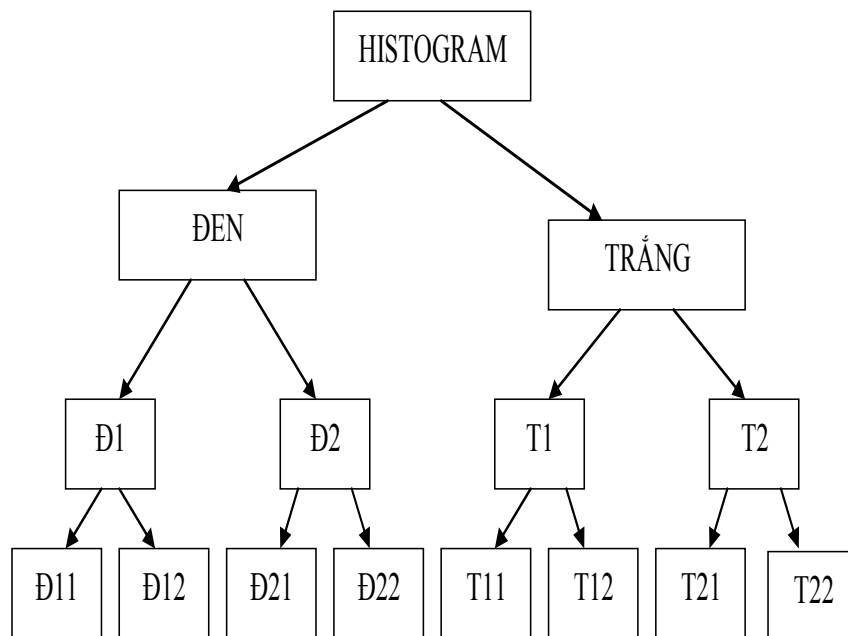
$$Y = 0.114 \cdot R + 0.587 \cdot G + 0.299 \cdot B. \quad (1.4)$$

Biểu đồ mức xám sẽ phân thành các loại ĐEN, TRẮNG. Với

$$\text{ĐEN} = \frac{\sum_{i=0}^{127} X_i}{\sum_{i=0}^{255} X_i}; \text{TRẮNG} = \frac{\sum_{i=128}^{255} X_i}{\sum_{i=0}^{255} X_i} \quad (1.5)$$

Trong đó, X_i là số pixel có mức xám i .

Rồi ĐEN sẽ phân thành Đ1, Đ2,... như sau:



Hình 1.7: Quá trình phân tách mức xám

Việc tạo chỉ mục sẽ được thực hiện tăng dần theo cột đen và giảm dần theo cột trắng theo 3 mức của cây nhị phân phía trên.

1.4.2 Truy vấn ảnh sử dụng chỉ mục màu sắc, hình dạng và vị trí [14]

B.G. Prasad, K.K. Biswas, và S.K. Gupta đã xây dựng hệ thống truy vấn ảnh dựa theo chỉ mục màu sắc hình dạng và vị trí dựa theo các vùng trên ảnh. Hệ thống

đưa ra chỉ mục kết hợp nhưng không tìm ra độ đo sự tương đồng dựa theo màu sắc hình dạng và vị trí mà chỉ dựa theo sự trùng khớp về chỉ mục.

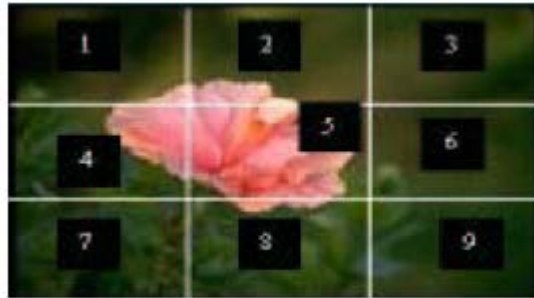
Hệ thống sẽ sử dụng không gian màu giảm 25 màu mà mắt người dễ phân biệt theo bảng tra cứu màu sau:

Bảng 1.2: Bảng thông số 25 màu cơ bản

Thứ tự	Màu	R	G	B
1	Black	0	0	0
2	Sea green	0	182	0
3	Light green	0	255	170
4	Olive green	36	73	0
5	Aqua	36	146	170
6	Bright green	36	255	0
7	Blue	73	36	170
8	Green	73	146	0
9	Turquoise	73	219	170
10	Brown	109	36	0
11	Blue gray	109	109	170
12	Lime	109	219	0
13	Lavender	146	0	170
14	Plum	146	109	0
15	Teal	146	182	170
16	Dark red	182	0	0
17	Magenta	182	73	170
18	Yellow green	182	182	0
19	Flouro green	182	255	170
20	Red	219	73	0
21	Rose	219	146	170
22	Yellow	219	255	0
23	Pink	255	36	170
24	Orange	255	146	0
25	White	255	255	255

Về hình dạng tiến hành phép quay chuẩn hóa trục và biến đổi tỷ lệ để chuẩn hóa kích thước và hướng. Sau đó tính toán các thông số hình dạng: trục chính, trục phụ, hình chữ nhật cơ sở, độ lệch tâm, trọng tâm, độ bao phủ dòng, và bao phủ cột.

Về vị trí sẽ chia ảnh làm 9 phân vùng vị trí đánh số như sau:



Hình 1.8: Phân vùng vị trí

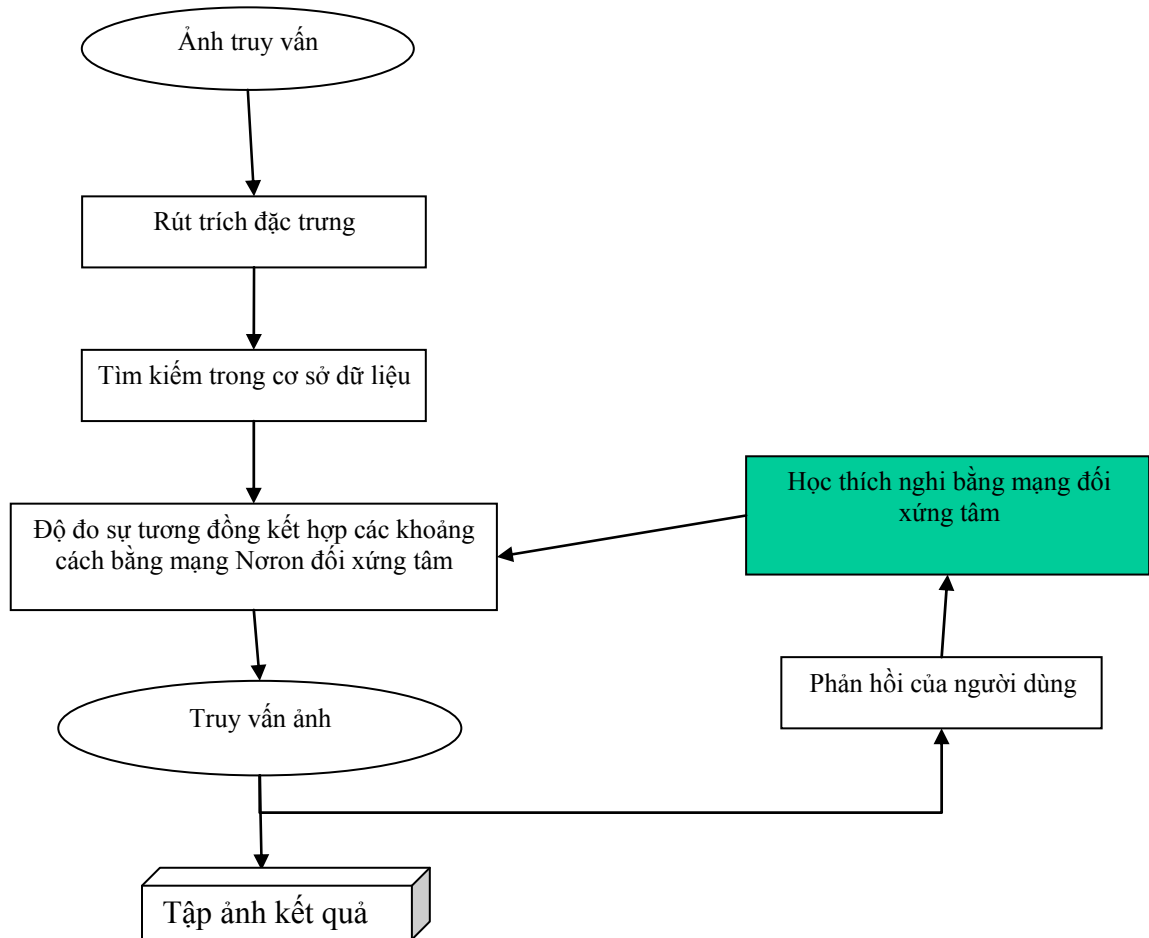
Kết quả thực nghiệm của hệ thống này như sau:

Bảng 1.3: Bảng kết quả thực nghiệm hệ thống của B.G. Prasad, K.K Biswas và S.K. Gupta

	Dựa trên màu sắc		Dựa trên hình dạng	
	Tỉ lệ precision	Tỷ lệ recall	Tỉ lệ precision	Tỷ lệ recall
Trung bình	0.743401	0.528	0.58338	0.393825
Độ lệch chuẩn	0.26772	0.218	0.205423	0.187294
	Dựa trên màu sắc- hình dạng		Dựa trên màu sắc – hình dạng – vị trí	
	Tỉ lệ precision	Tỷ lệ recall	Tỉ lệ precision	Tỷ lệ recall
Trung bình	0.932563	0.355601	0.788688	0.593624
Độ lệch chuẩn	0.176627	0.257311	0.279232	0.33318

1.4.3 Truy vấn ảnh dựa trên mạng Nơron [13]

HYOUNG K. LEE, SUK I. YOO đã xây dựng hệ thống Neural Network based Flexible Image Retrieval (NNFIR) sử dụng mạng Nơron đối xứng tâm (Radial Basis Function Network (RBFN)) kết hợp giá trị các đặc trưng để truy vấn ảnh.



Hình 1.9: Hệ thống truy vấn ảnh NNFIR

Rút trích các đặc trưng, mỗi ảnh q đều dùng không gian màu RGB, sẽ được đặc trưng bởi một vector đặc trưng F^q :

$$F^q = \{f_1^q, f_2^q, \dots, f_{29}^q\}$$

f_1^q, f_2^q, f_3^q là các histogram màu Red, Green, Blue xác định như sau:

$$f_i^q[k] = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \begin{cases} 1 & \text{với } q[i][x,y]=k \\ 0 & \text{ngược lại} \end{cases} \quad (1.6)$$

Với $0 \leq k \leq L$ và $i=1,2,3$ trong đó L số mức màu của ảnh, $M \times N$ là kích thước của ảnh. Màu mỗi pixel có tọa độ (x, y) là sự tổ hợp của 3 thành phần màu Red, Green, Blue; $q[1][x, y]$ là giá trị thành phần màu Red của pixel có tọa độ (x, y) của ảnh q , $q[2][x, y]$ là giá trị thành phần màu Green của pixel có tọa độ (x, y) của ảnh q , $q[3][x, y]$ là giá trị thành phần màu Blue của pixel có tọa độ (x, y) của ảnh q .

f_4^q, f_5^q, f_6^q và f_7^q, f_8^q, f_9^q lần lượt là trung bình và phương sai của các thành phần màu red, green, blue xác định như sau:

$$\begin{aligned} f_i^q &= \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} q[i-3][x,y] / MN & \text{khi } i = 4, 5, 6 \\ f_j^q &= \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (q[j-6][x,y] - f_{j-3}^q)^2 / MN & \text{khi } j = 7, 8, 9 \end{aligned} \quad (1.7)$$

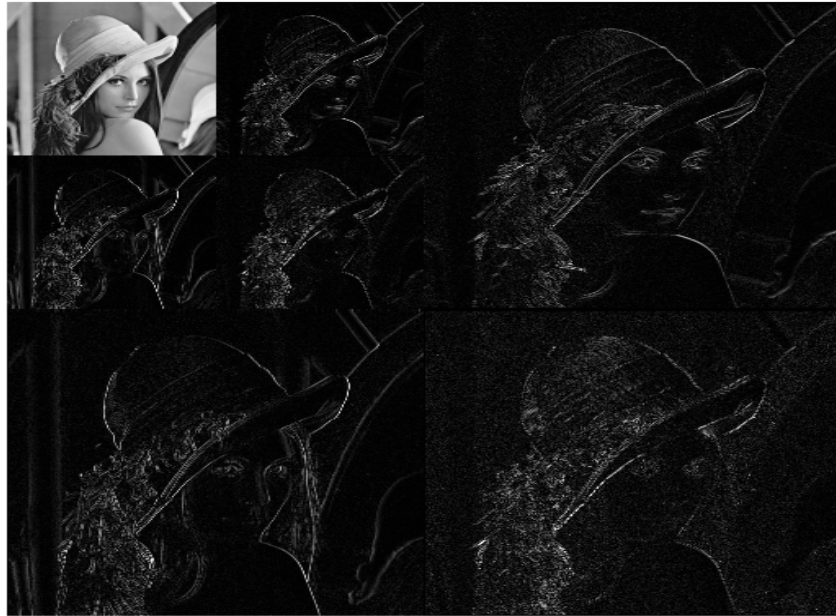
Sau đó áp dụng phép biến đổi wavelet song trực giao lên ảnh q đến mức 3 ta được các hệ số sau: xấp xỉ wavelet mức 1, wavelet mức 1 ngang, dọc, đường chéo, wavelet mức 2 ngang, dọc, đường chéo, wavelet mức 3 ngang, dọc, đường chéo. Ta có $f_{10}^q \dots f_{19}^q$ và $f_{20}^q \dots f_{29}^q$ là trung bình và phương sai của các hệ số wavelet thu được ở trên.



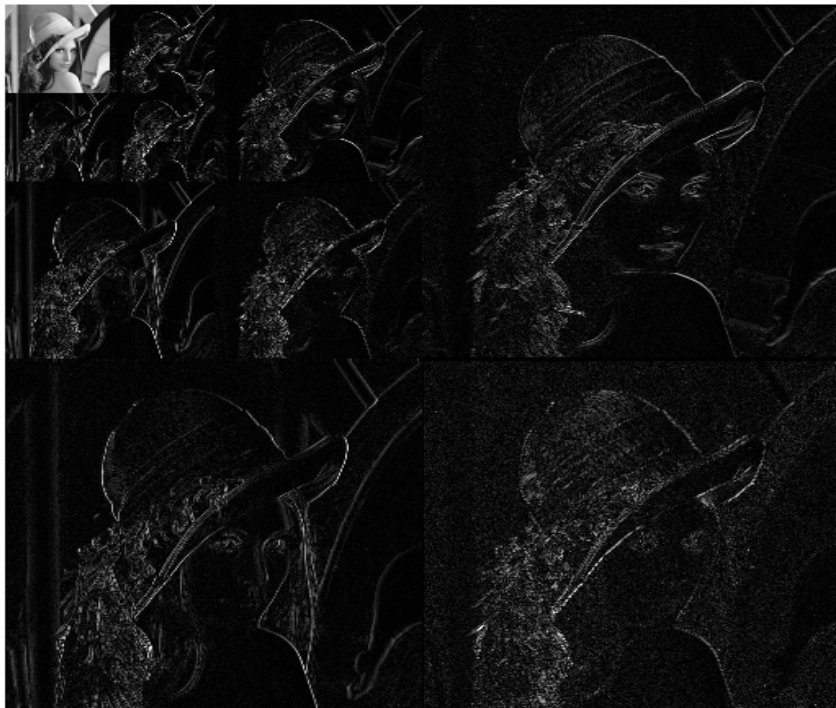
Hình 1.10: Ảnh gốc



Hình 1.11: Ảnh qua phép biến đổi wavelet song trực giao mức 1



Hình 1.12: Ảnh qua phép biến đổi wavelet song trực giao mức 2



Hình 1.13: Ảnh qua phép biến đổi wavelet song trực giao mức 3

Để so sánh histogram ta dùng độ đo phân giao của hai histogram như sau:

$$v_i = 1 - \frac{\sum_{m=0}^{M-1} \min(f_i^q[m], f_i^t[m])}{\min(f_i^q, f_i^t)} \quad (1.8)$$

$$\text{với } |f_i^q| = \sum_{m=0}^{L-1} f_i^q, |f_i^t| = \sum_{m=0}^{L-1} f_i^t$$

Với $i=1,2,3$. Để so sánh các đặc trưng khác ta sử dụng phép trừ đơn giản:

$$v_i = f_i^q - f_i^t, \quad i=4,5,...,29 \quad (1.9)$$

Vai trò của mạng Noron RBF: dùng để kết hợp các giá trị khoảng cách v_i . Input của mạng Noron RBF là tập các khoảng cách v_k giữa 2 ảnh I_1, I_2 . Nếu tất cả các đặc trưng của 2 ảnh này giống nhau ta mong muốn Output là 1 và nếu hoàn toàn khác nhau ta mong muốn Output là 0. Khi nhận được tập ảnh kết quả của ảnh truy vấn thì người dùng có thể phản hồi trên từng ảnh mà mình thấy là giống nhau (alike), tương tự (similar) hay khác nhau (different). Hàm RBF là

$$f(\bar{x}) = \sum_{i=1}^R c_i g(\|\bar{x} - \bar{x}_i\|) \quad (1.10)$$

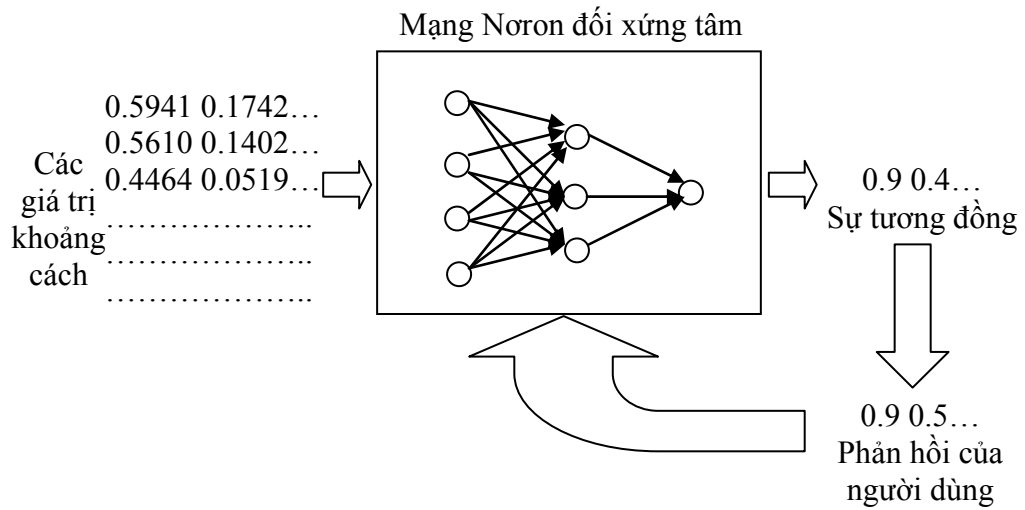
với g là một hàm Gauss có dạng:

$$g(\|\bar{x} - \bar{x}_i\|) = \exp\left(-\frac{\|\bar{x} - \bar{x}_i\|^2}{2\sigma^2}\right) \quad (1.11)$$

Giá trị phản hồi của người dùng về kết quả truy vấn xác định như sau:

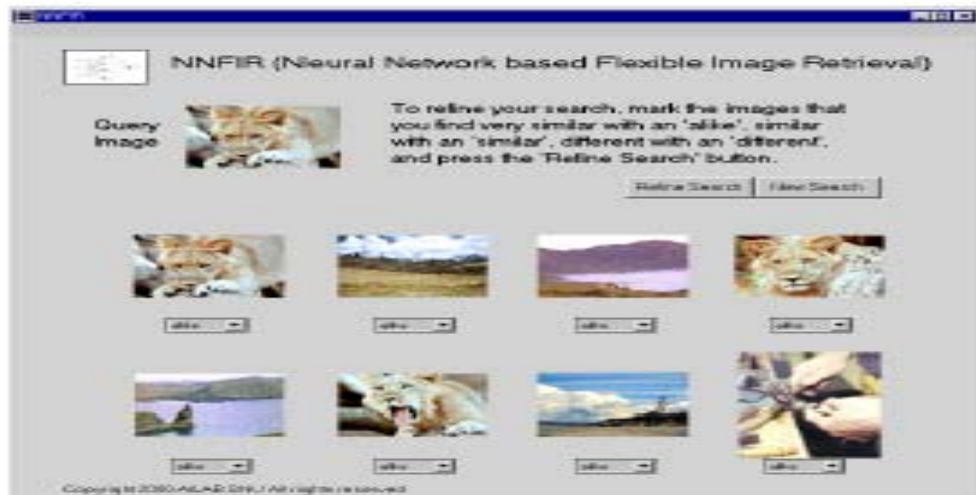
$$UF(q) = \begin{cases} 0.9 & \text{nếu giống nhau} \\ 0.5 & \text{nếu tương tự} \\ 0 & \text{nếu khác nhau} \end{cases} \quad (1.12)$$

Mạng Noron RBF có thể minh họa như sau:



Hình 1.14: Mối quan hệ giữa khoảng cách, RBNF và UF

Quá trình huấn luyện sẽ điều chỉnh các tham số c_i dựa theo phản hồi của người dùng. Từ đó cho kết quả truy vấn ngày càng chính xác hơn. Hệ thống cho kết quả khả quan sau 5 lần phản hồi của người dùng.



Hình 1.15: Kết quả lần truy vấn lần đầu



Hình 1.16: Kết quả truy vấn sau khi nhận phản hồi

Kết quả đạt được của hệ thống NNFIR so với các cách kết hợp tuyến tính như sau:

Bảng 1.4: Kết quả hệ thống NNFIR

	Phương pháp kết hợp tuyến tính	NNFIR
Tỷ lệ bao phủ cho cửa sổ kích thước W=8	71.32%	81.70%
Tỷ lệ chính xác khi bao phủ là 100%	40.11%	54.22%
Thời gian kết hợp	7.33 msec	16.92 msec

Độ bao phủ (Recall) = Số ảnh đúng trả ra/Tổng số ảnh đúng trong CSDL.

Độ chính xác (Precision) = Số ảnh đúng trả ra/Tổng số ảnh trả ra.

Chương 2: ĐỀ XUẤT HỆ THỐNG LIÊN MẠNG META-NƠRON

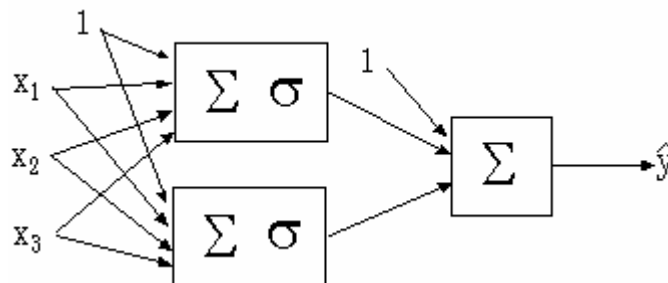
Chương này trình bày một hệ thống liên mạng meta-Nơron dùng kết hợp nhiều mạng Nơron với nhau áp dụng cho bài toán truy vấn ảnh tổng quát. Hệ thống đề xuất này dựa trên sự cải tiến phương pháp kết hợp nhiều mạng Nơron lại với nhau bằng thuật giải di truyền hay Logic mờ của các tác giả trước đây.

2.1 Mạng Nơron

Mạng Nơron xuất phát từ việc mô phỏng hoạt động của bộ não người bằng máy tính. Mạng Nơron là một mô hình tính toán chứa các đơn vị xử lý có khả năng truyền thông với nhau thông qua các liên kết có trọng số và có khả năng học từ kinh nghiệm hay từ tập mẫu. Vì vậy mạng Nơron được xem như là một hàm số với tham số có thể điều chỉnh cho thích hợp qua quá trình huấn luyện.

Cấu trúc của một mạng Nơron bao gồm một số lớp: 1 lớp input, 1 hay nhiều lớp ẩn và 1 lớp output. Mỗi lớp bao gồm một số các đơn vị hay nút. Đầu ra của mỗi đơn vị được liên kết thông qua trọng số đến đơn vị khác ở lớp kế tiếp.

Mạng Nơron là một cấu trúc bao hàm các quan hệ các trọng số của Nơron, hay các đơn vị (units) với các phép biến đổi vô hướng không tuyến tính và tuyến tính.



Hình 2.1: Mạng Nơron truyền thẳng

Hình 2.1 trên là một ví dụ minh họa về mạng Nơron gồm 3 lớp: 1 lớp input, 1 lớp ẩn và 1 lớp output. Lớp input có 3 nút nhập, $x = \{x_1, x_2, x_3\}$ và đầu vào ngoài dữ liệu (độ lệch) nối vào 2 Nơron trong lớp ẩn. Hai giá trị ra từ lớp ẩn cùng với độ lệch được đưa vào 1 Nơron ở lớp xuất tính ra giá trị xuất vô hướng \hat{y} .

Lớp chứa các Nơron gọi là lớp ẩn vì giá trị xuất của nó không cần xuất ra trực tiếp dữ liệu. Các mũi tên tượng trưng cho các tham số giá trị thực hay trọng số của mạng.

Giá trị xuất của mạng có L nút xuất là:

$$\hat{y}_i = g(v_i), \quad i=1, \dots, K. \quad (2.1)$$

Trong đó $g(v_i)$ là hàm truyền theo v_i

$$v_i = b_{0i} + \sum_{j=1}^J b_{ji} h_j \quad (2.2)$$

có J nút ẩn với kết xuất h_j , b_{ji} là các trọng trên cung liên kết từ nút ẩn j đến nút xuất thứ i, còn b_{0i} là trọng ngưỡng của nút xuất thứ i.

Các công thức trên giải thích theo thuật ngữ mạng như sau. Mỗi nút trong J nút ẩn chuyển giá trị kết xuất h_j của nó đến từng nút xuất. Mỗi nút trong L nút xuất tính tổng trọng hóa theo J kết xuất của nút ẩn và cộng với ngưỡng b_{0i} . Kết xuất \hat{y}_i là giá trị hàm truyền của tổng này.

Mục tiêu của việc huấn luyện mạng là tìm giá trị trọng số thích hợp với dữ liệu nhập x bất kỳ vào mạng sẽ cho ra giá trị output xấp xỉ output mong đợi. Các thuật toán huấn luyện sẽ lặp nhiều bước với trọng số khởi tạo và được cập nhật tốt hơn trong mỗi bước lặp.

Trước khi mạng được chấp nhận, cần có bước kiểm nghiệm (validate). Nghĩa là mạng sẽ được chạy thử với một số dữ liệu đã biết. Đơn giản và tốt nhất là thử nghiệm mạng với tập dữ liệu mà không có trong dữ liệu được huấn luyện và dữ liệu này với dữ liệu huấn luyện được tạo ra trong cùng các điều kiện tương tự nhau. Việc huấn luyện mạng nếu thất bại trong bước kiểm nghiệm, trong trường hợp này, ta phải chọn mô hình tốt hơn. Tuy nhiên, đôi khi ta chỉ cần lặp lại bước huấn luyện với giá trị tham số khởi tạo khác.

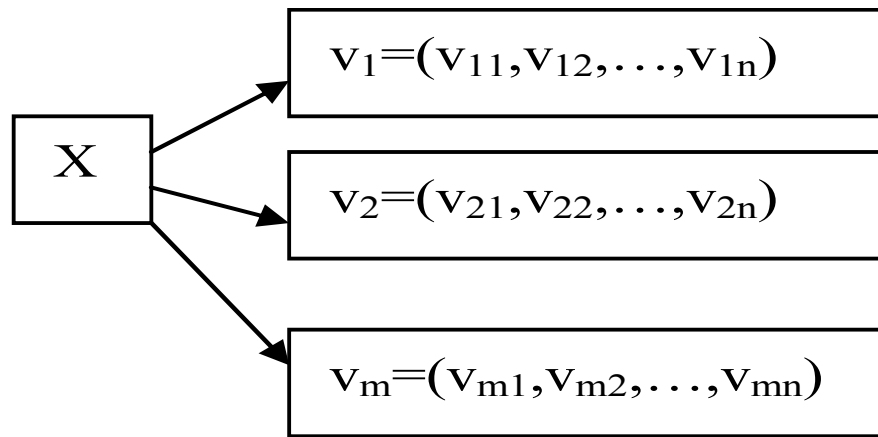
Mỗi bài toán mạng Noron thường thực hiện qua 4 bước sau:

- Chuẩn bị tập dữ liệu nhập.
- Khởi tạo mạng.
- Huấn luyện mạng.
- Kiểm định mô hình.

2.2 Hệ thống liên mạng meta-Noron

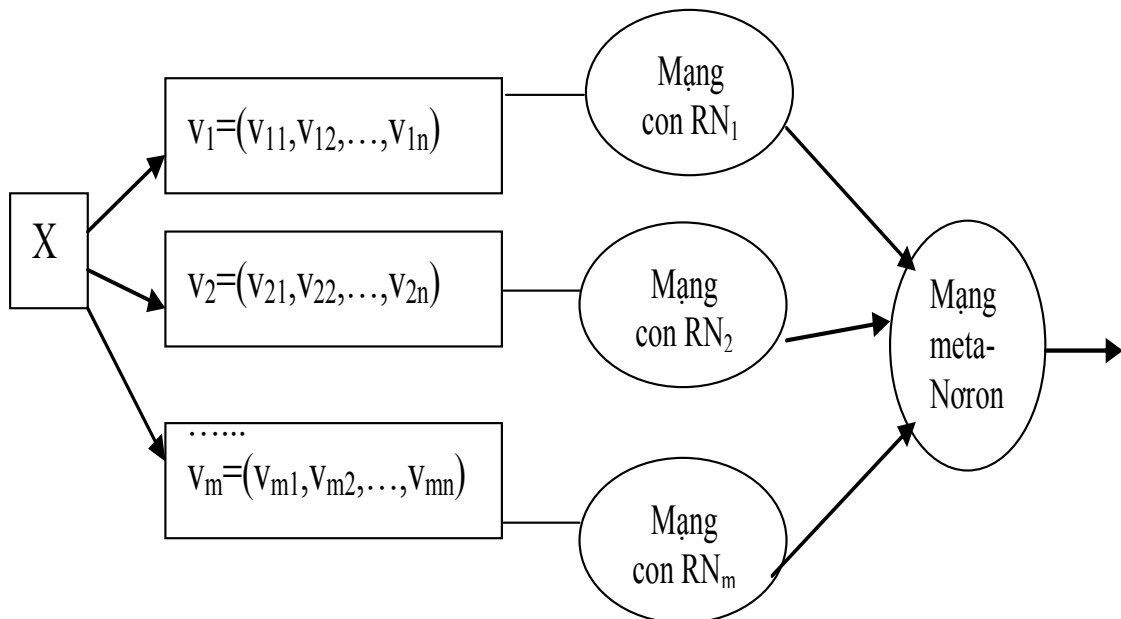
2.2.1 Giới thiệu các phương pháp kết hợp

Xét mẫu X gồm m vector đặc trưng (X được xem xét dựa trên m tiêu chí khác nhau) $X=(v_1, v_2, \dots, v_m)$. Mỗi vector v_k của mẫu X có n chiều $v_k=(v_{k1}, v_{k2}, \dots, v_{kn})$ với k từ 1 đến m .



Hình 2.2: Mẫu X đặc trưng bởi m vector

Xét phân hoạch $\Omega = \{\Omega_i \mid 1 \leq i \leq L\}$. Hai mẫu X^1, X^2 sẽ thuộc cùng phân hoạch nếu kết quả đánh giá hai mẫu này dựa trên m vector đặc trưng (hay m tiêu chí) này gần nhau theo công thức độ đo sự tương đồng 2.5. Xây dựng một hệ thống liên mạng meta-Nơron để đánh giá mẫu X gồm m vector đặc trưng như sau:

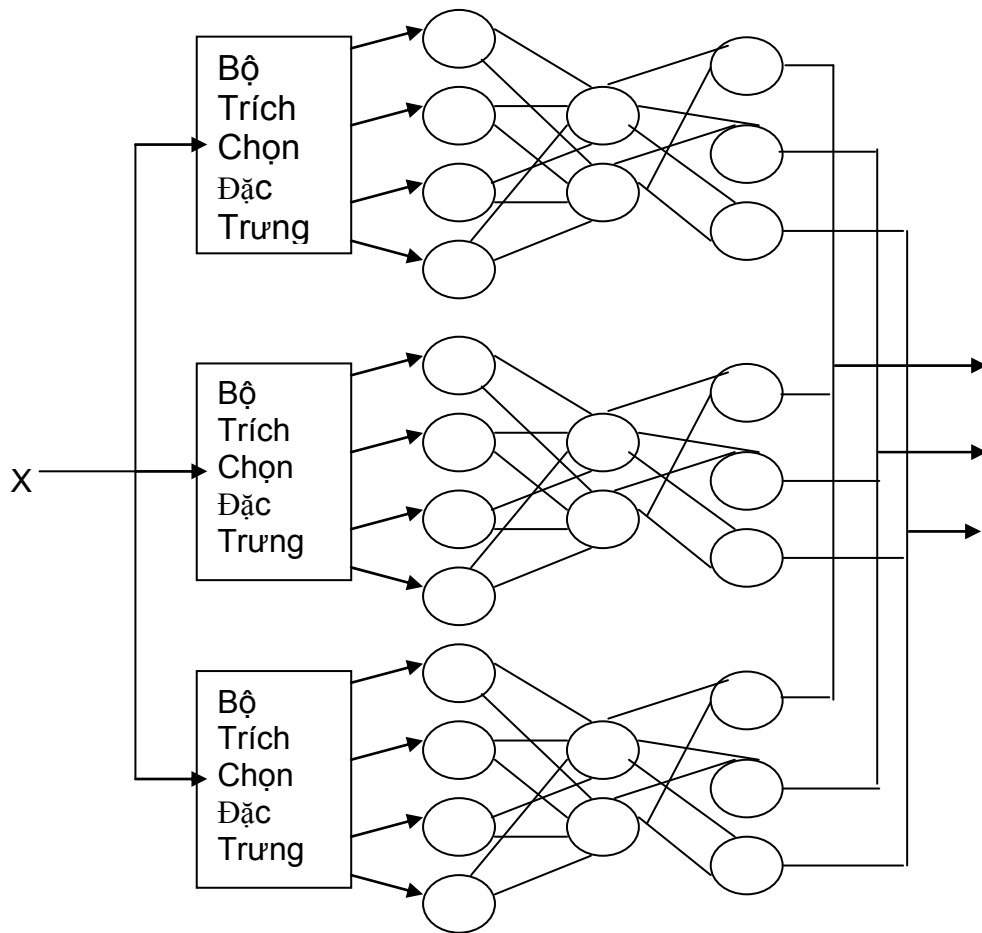


Hình 2.3: Tổng thể hệ thống liên mạng

Mỗi mạng Nơron con thứ k sẽ dùng để đánh giá thành phần v_k (hay tiêu chí thứ k) của mẫu X . Với từng mạng con phải tạo ra tập huấn luyện và tập mẫu kiểm

tra để tiến hành huấn luyện nhằm xác định bộ trọng số tối ưu cho mạng con. Sau đó, mỗi mẫu X qua m mạng con sẽ có được m kết luận đánh giá về mẫu X .

Như vậy mỗi vector đặc trưng v_k qua mạng con thứ k sẽ cho ta một vector xác suất P_k cho ta biết xác suất X thuộc về phân hoạch thứ i nào với i từ 1 đến L . Căn cứ theo một tiêu chí thứ k ta cũng có thể cho ra một quyết định X thuộc về phân hoạch nào. Giả sử $P_k(\Omega_i | X) = 0.9$ thì nghĩa là dựa theo tiêu chí thứ k (vector đặc trưng v_k) thì mạng con cho rằng 90% là mẫu X thuộc về phân hoạch thứ i .

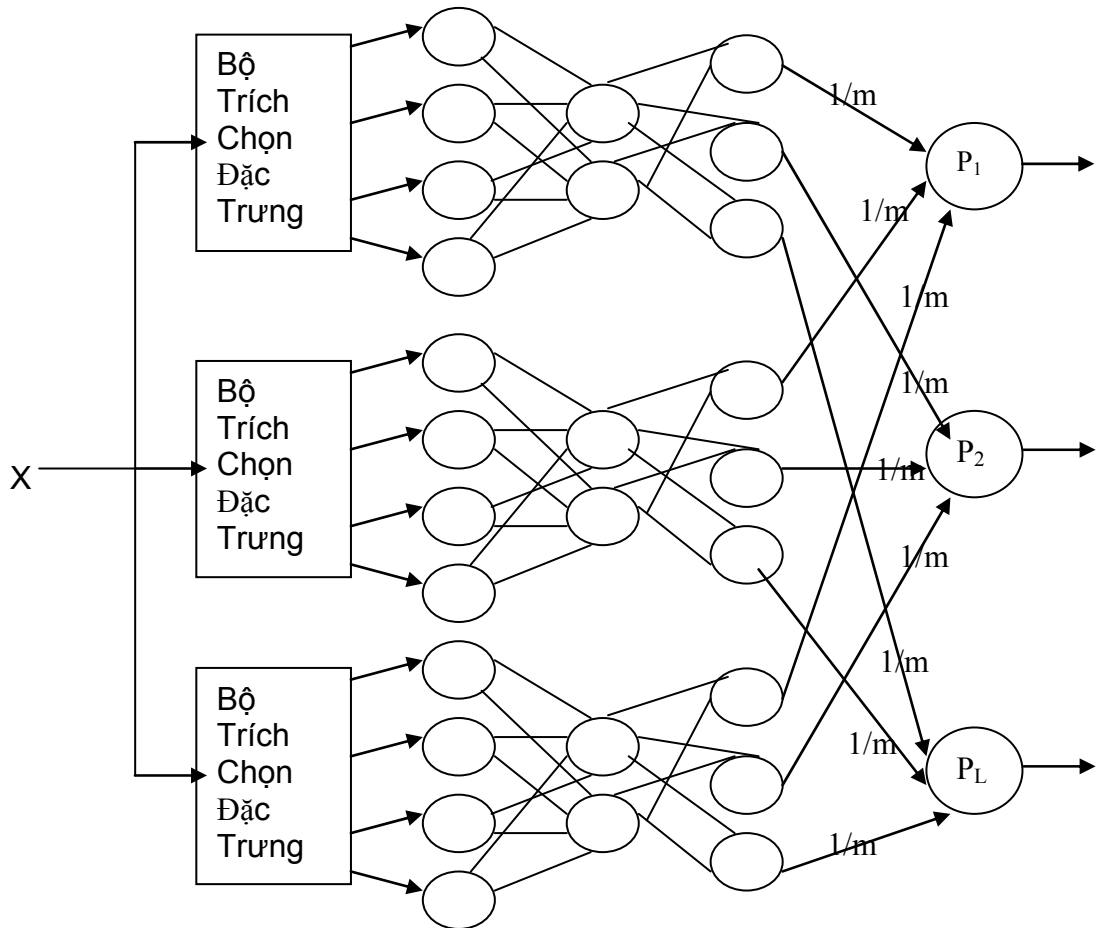


Hình 2.4: Quá trình xử lý của các mạng con

Nhất thiết phải hợp nhất các đánh giá của các mạng con để cho ra kết luận cuối cùng của toàn hệ thống [5]. Giả sử ta đánh đồng các kết luận của các mạng con

(hay xem các tiêu chí đánh giá X là như nhau) thì kết luận cuối cùng của toàn hệ thống sẽ là:

$$P(\Omega_i | X) = \sum_{k=1}^m \frac{1}{m} P_k(\Omega_i | X) \quad (2.3)$$



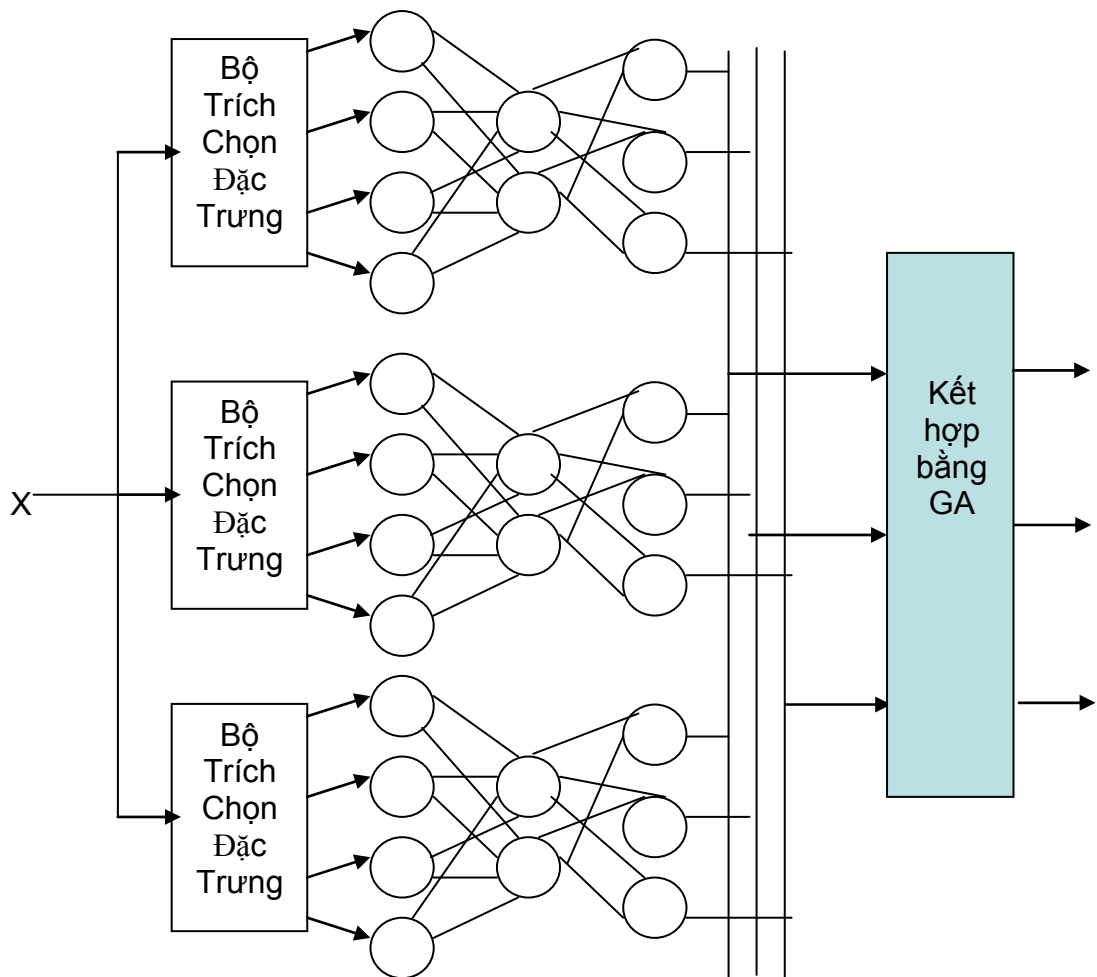
Hình 2.5: Phương pháp kết hợp trung bình kết quả các mạng con

Phương pháp kết hợp trung bình là phương pháp kết hợp đơn giản, tốn ít chi phí tính toán và nhanh chóng cho ra kết quả tổng hợp của toàn hệ thống.

Tuy nhiên trong thực tế mức độ quan trọng của các tiêu chí là khác nhau nên cần xác định hệ số tin cậy r_j^i tương ứng. Có thể dùng các kỹ thuật khác như thuật

giải di truyền để xác định bộ hệ số tin cậy tối ưu. Kết luận cuối cùng của hệ thống sẽ thành [5]:

$$P(\Omega_i | X) = \sum_{k=1}^m r_j^i P_k(\Omega_i | X) \quad (2.4)$$



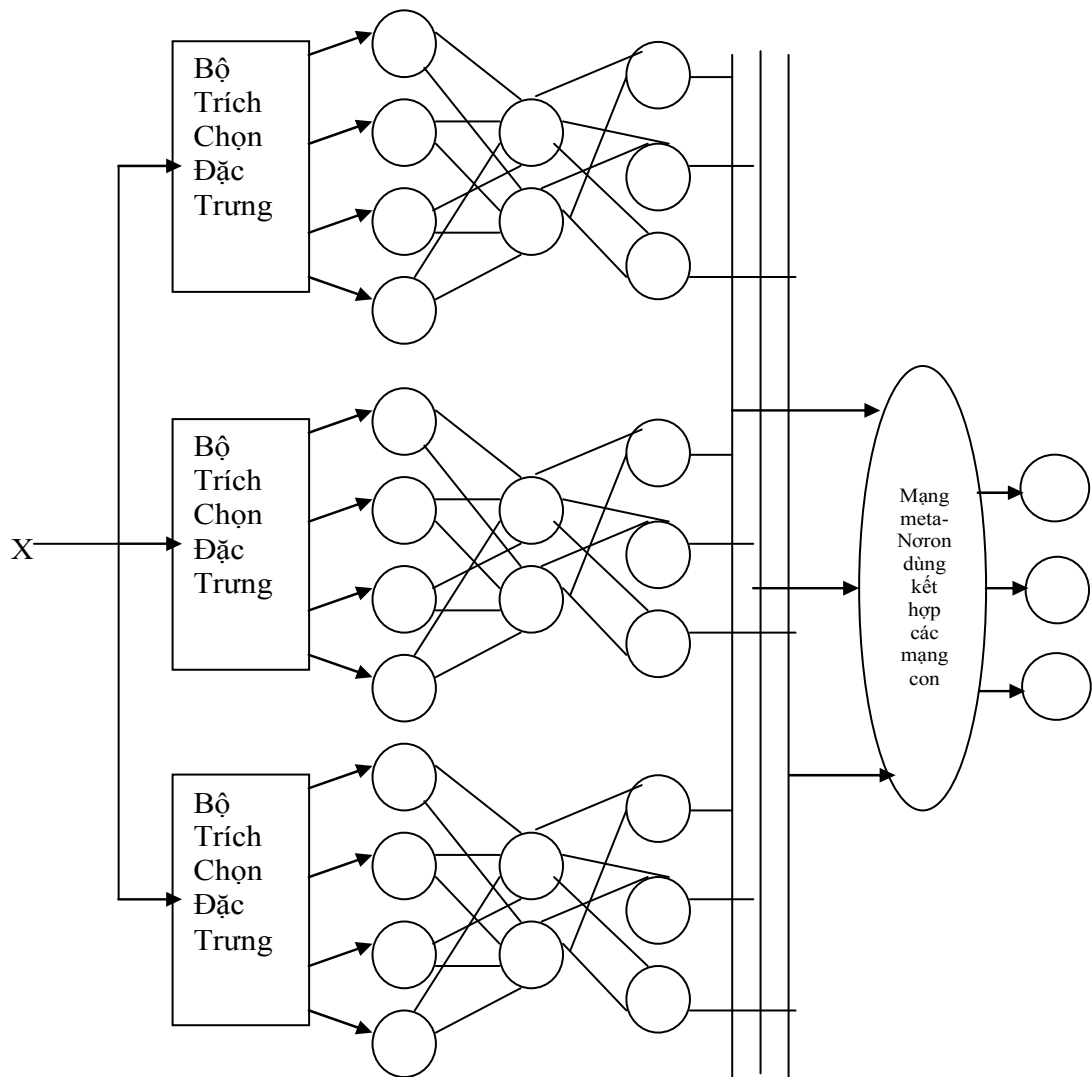
Hình 2.6: Kết hợp mạng Nơron và thuật giải di truyền (NN_GA)

Đây là dạng kết hợp tuyến tính các kết quả của các mạng con. Để sự kết hợp này thêm uyển chuyển linh hoạt tùy vào bài toán, các kết quả của các mạng con sẽ được cho qua mạng meta-Nơron toàn cục thích hợp để cho ra kết luận cuối cùng.

Có thể áp dụng mạng Noron truyền thẳng với hàm tuyến tính hay mạng Noron phi tuyến vào mạng meta-Noron toàn cục để được một sự kết hợp output các mạng con thích hợp tùy từng trường hợp cụ thể. Bộ trọng số của mạng meta-Noron sẽ được xác định qua quá trình “học”.

Vì vậy cần có tập mẫu huấn luyện cho mạng meta-Noron toàn cục. Nếu như input của tập mẫu huấn luyện của các mạng con là vector đặc trưng tương ứng v_k thì input tập mẫu huấn luyện của mạng meta-Noron toàn cục là output của các mạng con.

Như vậy, hệ thống liên mạng meta-Noron này sẽ kết hợp các mạng Noron con một cách linh hoạt tùy thuộc vào việc lựa chọn cấu hình của mạng meta-Noron để có sự kết hợp phù hợp với từng lớp bài toán cụ thể. Ngoài ra, hệ thống liên mạng meta-Noron không đòi hỏi các nhà phát triển, ứng dụng hệ thống meta-Noron phải nắm rõ nhiều loại kỹ thuật khó khác nhau như Noron và thuật giải di truyền (GA) hay SVM... mà chỉ cần hiểu rõ kỹ thuật về mạng Noron thuần túy.



Hình 2.7: Tổng quan mô hình kết hợp bằng hệ thống liên mạng

Tư tưởng chính của hệ thống liên mạng meta-Nơon là thay vì dùng thuật giải di truyền kết hợp các output của các mạng con thì hệ thống sử dụng ngay chính công cụ mạng Nơon để kết hợp các output của các mạng con. Các output của các mạng con sẽ được kết hợp bằng mạng meta-Nơon toàn cục. Mạng meta-Nơon toàn cục **gồm nhiều mạng Nơon** thành phần. Chi tiết về mạng meta-Nơon toàn cục cũng như cách kết hợp các output của các mạng con vào mạng meta-Nơon toàn cục sẽ trình bày trong phần sau 2.2.2.

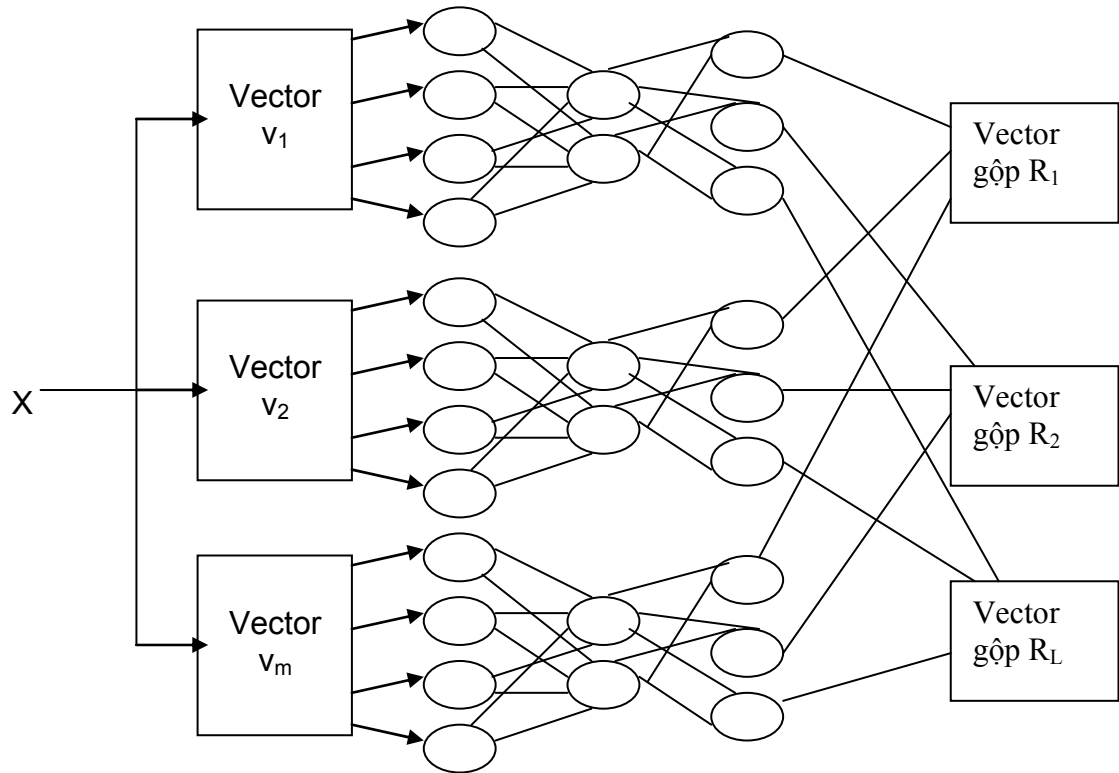
Để thấy rõ hơn lý do của việc phải kết hợp các kết quả của các mạng con, trong hệ thống thử nghiệm truy tìm vùng du lịch với 822 ảnh huấn luyện và 82 ảnh dùng để kiểm tra, luận văn đã thử xác định vùng du lịch bằng các phương pháp: căn cứ theo từng mạng con, theo phương pháp trung bình và phương pháp dùng hệ thống liên mạng kết quả các mạng con (xem chi tiết tại bảng 4.1) thấy rằng nếu không kết hợp các kết quả của các mạng con thì kết quả sẽ không cao. Vì mỗi mạng con chỉ căn cứ trên một tiêu chí để ra quyết định. Trong khi thực tế thì mỗi tiêu chí sẽ có một hệ số tin cậy khác nhau. Ví dụ, để đánh giá khả năng tư duy logic thì kết quả các môn tự nhiên sẽ có hệ số tin cậy cao hơn các môn xã hội. Các công trình của các tác giả trước đây thường kết hợp kết quả bằng thuật giải di truyền hay Logic mờ hay SVM... Trong luận văn này sử dụng mạng meta-Nơron toàn cục để kết hợp kết quả đánh giá của các mạng con cho ra kết quả đánh giá cuối cùng.

2.2.2 Kiến trúc của hệ thống liên mạng meta-Nơron

Giả sử cần đánh giá mẫu X (được đặc trưng bởi m vector, mỗi vector n chiều) thuộc về phân hoạch nào trong $1, 2, \dots, L$ phân hoạch. Như vậy, ta cần m mạng Nơron con để đánh giá xem X thuộc phân hoạch nào theo tiêu chí (hay vector đặc trưng) tương ứng. Mỗi mạng con sẽ có n nút input (số chiều của vector đặc trưng) và có L nút output (số phân hoạch).

Mạng có thể có nhiều lớp ẩn thay vì chỉ 1 lớp ẩn. Có nhiều nghiên cứu công phu nhằm tìm hiểu xem việc thêm các lớp ẩn có làm tăng năng lực học hoặc tốc độ học của mạng như việc thêm số nút ẩn không. Tuy nhiên, chưa có lập luận lý thuyết nào ủng hộ việc thêm các lớp ẩn có tính thuyết phục. Ngược lại, mạng có một lớp ẩn có thể ánh xạ được bất kỳ ánh xạ nào. Nhiều lớp ẩn gây ra bất lợi vì làm cho các công thức toán học của mạng phức tạp hơn và các chương trình máy tính cài đặt theo các công thức này cũng khó hơn. Vì thế, luận văn này đề xuất sử dụng mạng Nơron chỉ có 1 lớp ẩn cho các mạng Nơron trong hệ thống liên mạng meta-Nơron.

Mạng Noron con thứ k ký hiệu là RN_k . Output thứ j (j từ 1 đến L) của mạng con thứ k (k từ 1 đến m) ký hiệu là ro_{kj} , cho ta xác suất đánh giá mẫu X thuộc về phân hoạch thứ j theo đánh giá của mạng con thứ k .



Hình 2.8: Gộp kết quả các mạng con chuẩn bị đưa vào mạng toàn cục

Gộp các output thứ nhất của tất cả các mạng con làm thành phần của vector $R_1(ro_{11}, ro_{21}, \dots, ro_{m1})$, vector R_1 này sẽ chứa tất cả các đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=1$ của tất cả m mạng con.

Tương tự, gộp các output thứ k của tất cả các mạng con làm thành phần của vector $R_k(ro_{1k}, ro_{2k}, \dots, ro_{mk})$, vector R_k này sẽ chứa tất cả các đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=k$ của tất cả m mạng con.

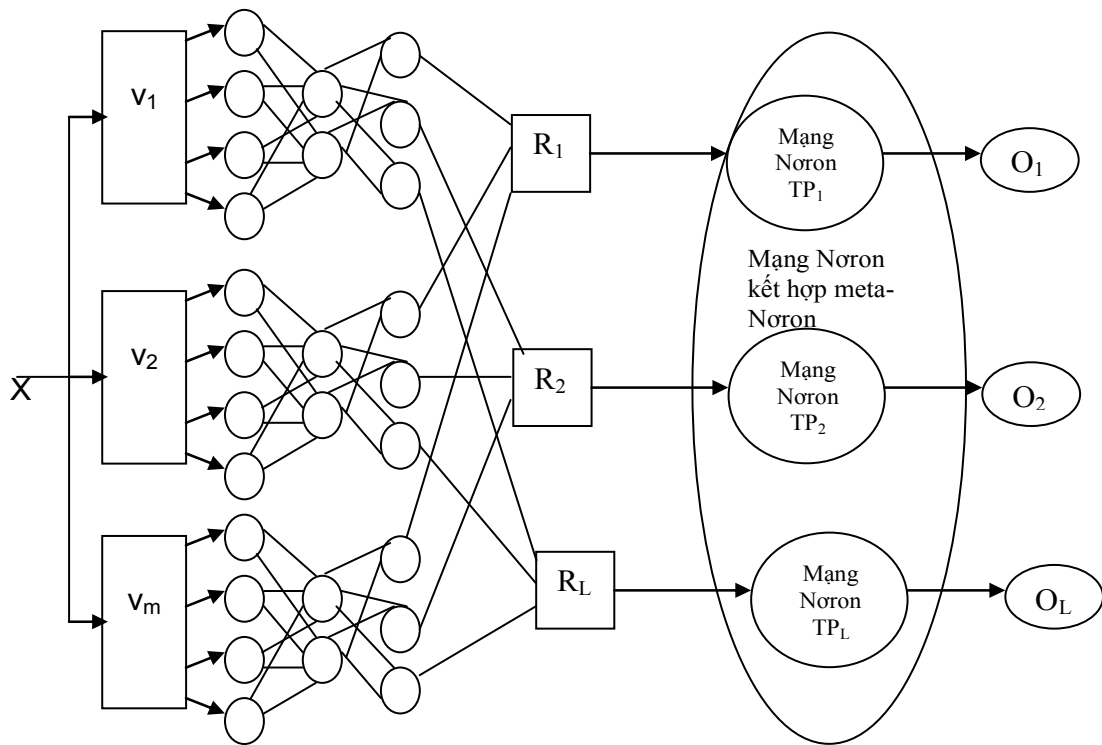
Gộp các output thứ L của tất cả các mạng con làm thành phần của vector $R_L(ro_{1L}, ro_{2L}, \dots, ro_{mL})$, vector R_L này sẽ chứa tất cả các đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=L$ của tất cả m mạng con.

Vector ra R_1 sẽ là input của một Noron thành phần ký hiệu TP_1 để cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất. Như vậy mạng Noron thành phần TP_1 sẽ có m nút input (số chiều của vector ra R_1) và một nút output O_1 cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=1$ căn cứ theo đánh giá của tất cả các mạng con.

Vector ra R_k sẽ là input của một Noron thành phần ký hiệu TP_k để cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ k . Như vậy mạng Noron thành phần TP_k sẽ có m nút input (số chiều của vector ra R_k) và một nút output O_k cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=k$ căn cứ theo đánh giá của tất cả các mạng con..

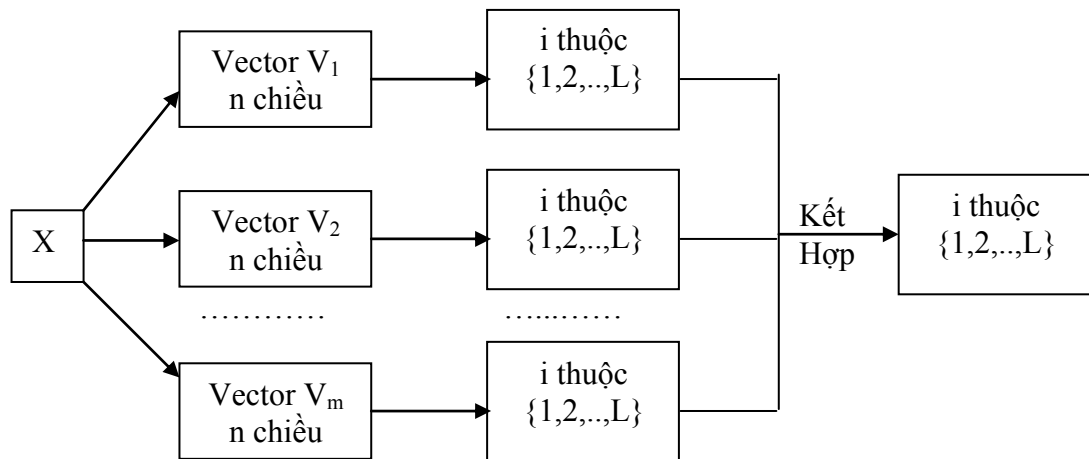
Vector ra R_L sẽ là input của một Noron thành phần ký hiệu TP_L để cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ L . Như vậy mạng Noron thành phần TP_L sẽ có m nút input (số chiều của vector ra R_L) và một nút output O_L cho ra đánh giá xác suất mẫu X thuộc về phân hoạch thứ nhất $i=L$ căn cứ theo đánh giá của tất cả các mạng con..

Có tất cả L vector ra R_L nên sẽ cần L mạng Noron thành phần TP_L . Mỗi mạng Noron thành phần có m nút input và một nút output duy nhất.



Hình 2.9: Đưa các vector R_i vào mạng kết hợp

Tóm lại, để phân hoạch mẫu X (đặc trưng bởi m vector, mỗi vector n chiều) vào một trong L phân hoạch $1, 2, \dots, L$ theo dạng sơ đồ sau [5],[6]:

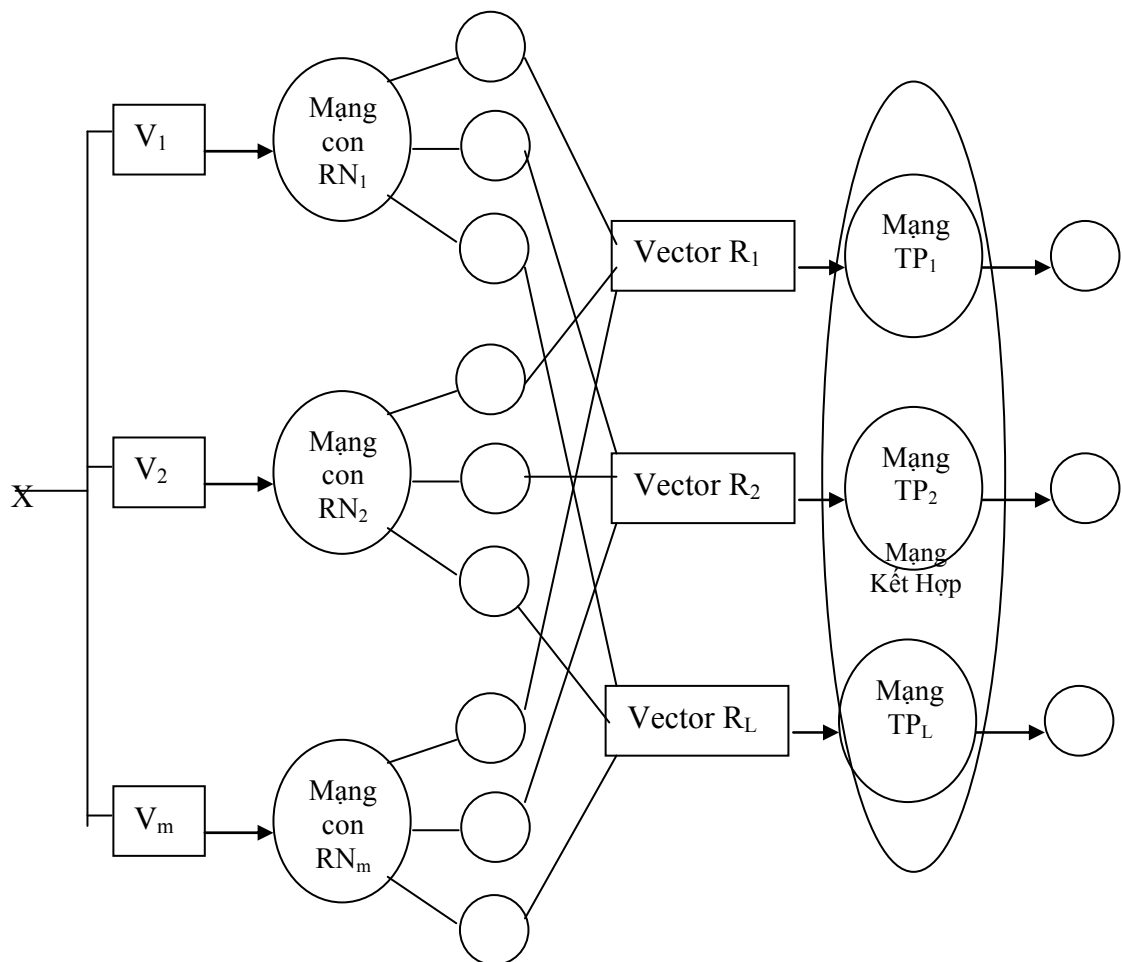


Hình 2.10: Phân lớp mẫu m vector

Hệ thống liên mạng cần xây dựng sẽ có m mạng Nơron con RN_k với k từ 1 đến m và một mạng meta-Nơron toàn cục gồm L mạng Nơron thành phần TP_i với i

từ 1 đến L . Trong đó các mạng Noron con cũng như mạng Noron thành phần của mạng toàn cục đều có 3 lớp: 1 lớp input, 1 lớp output và 1 lớp ẩn. Vì chưa có công thức xác định số nút ẩn tối ưu cho mạng Noron nên số nút ẩn tối ưu cần xác định qua quá trình thực nghiệm.

Cụ thể các mạng Noron con RN_k sẽ có n nút input (số chiều của vector đặc trưng) và có L nút output (số phân hoạch). Các mạng Noron thành phần TP_i của mạng toàn cục có m nút input (số mạng con) và có một nút output chính là nút output thứ i của toàn hệ thống liên mạng meta-Noron.



Hình 2.11: Mô hình liên mạng (m, n, L)

Trong đó, m là số vector đặc trưng của một mẫu, n là số chiều của các vector đặc trưng và L là số phân hoạch. Các mạng con RN_k với k từ 1 đến m sẽ có n nút input và L nút output. Các mạng thành phần TP_i với i từ 1 đến L có m nút input và 1 nút output duy nhất.

Sự kết hợp của hệ thống liên mạng là sự kết hợp output của các mạng con. Cụ thể bộ trọng số của các mạng thành phần TP_i sẽ kết hợp tất cả các output thứ i của các mạng con cho ra kết xuất thứ i của toàn hệ thống liên mạng.

2.2.3 Quá trình huấn luyện của hệ thống liên mạng

Quá trình huấn luyện cho hệ thống liên mạng theo 2 giai đoạn và thuộc dạng học có giám sát:

- Huấn luyện lần lượt cho tất cả các mạng con.
- Huấn luyện cho mạng meta-Noron toàn cục.

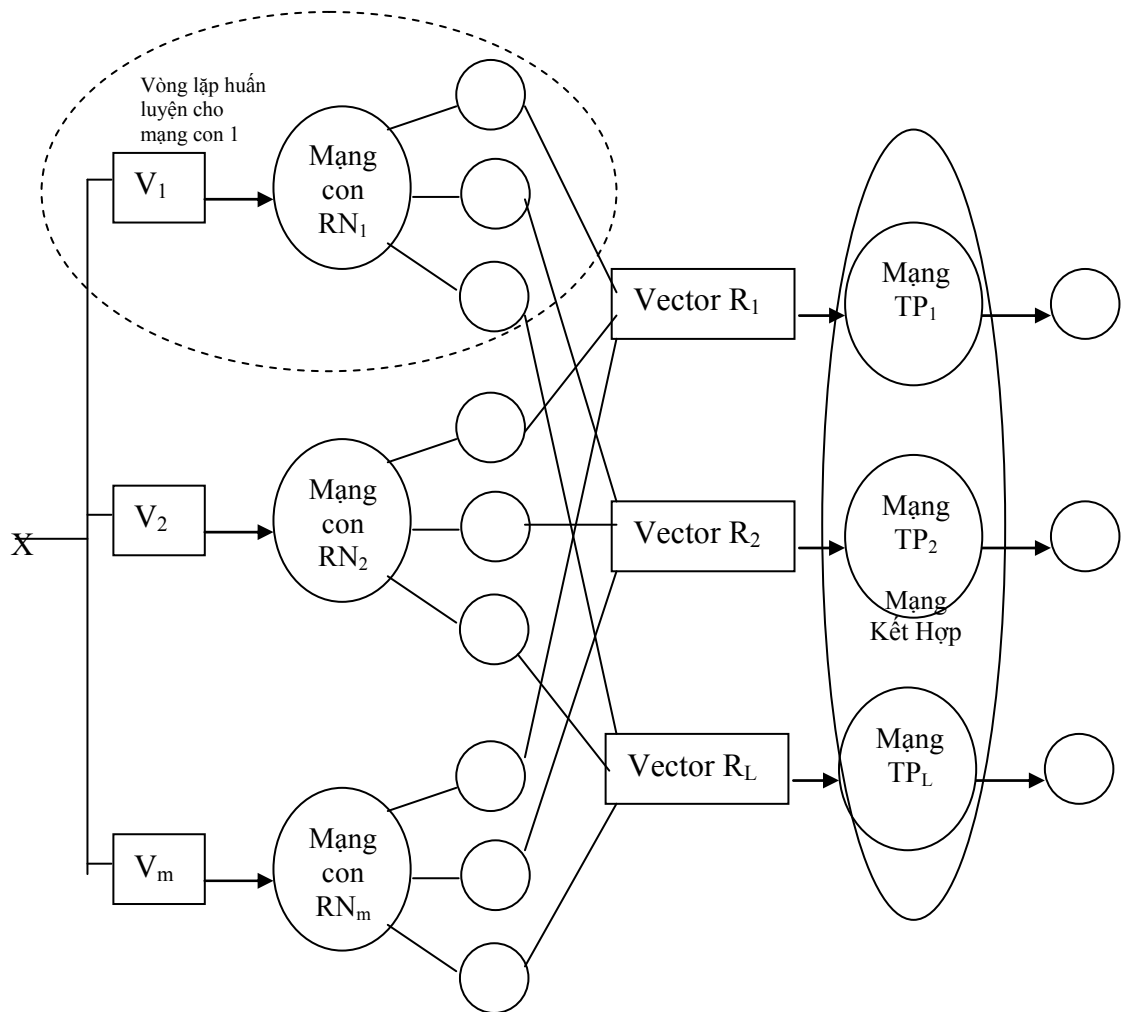
Xét những mẫu X đặc trưng bởi m vector v_k , mỗi vector n chiều, thuộc một trong L phân hoạch Ω_i với i từ 1 đến L . Giả sử tập mẫu X có tổng cộng TX mẫu đã biết trước phân hoạch của chúng. Nếu X thuộc về phân hoạch thứ $i=1$ thì vector output kỳ vọng (mong muốn) của X qua hệ thống liên mạng là $(1, 0, 0, \dots, 0)$ nghĩa là thành phần thứ nhất của vector output là 1, các thành phần còn lại của vector output là 0. Quá trình học là quá trình điều chỉnh bộ trọng số để output gần với kết quả kỳ vọng.

2.2.3.1 Huấn luyện cho các mạng con

Có TX mẫu X đã biết output, mỗi mẫu X rút ra được m vector đặc trưng v_k với k từ 1 đến m .

Vậy ta có TX vector v_1 đã biết output sẽ là tập dữ liệu huấn luyện cho mạng Noron con RN_1 . Nghĩa là lần lượt các vector v_1 này sẽ qua mạng con RN_1 tính ra

các output và điều chỉnh bộ trọng số của mạng Noron con RN_1 sao cho output của nó gần về output đã biết theo thuật toán học của mạng Noron. Lặp đi lặp lại quá trình điều chỉnh bộ trọng số cho đến khi hệ số lỗi chấp nhận được hay vượt quá số bước giới hạn. Sau khi huấn luyện xong cho mạng con RN_1 ta ghi nhận lại bộ trọng số của mạng con này. Chi tiết về thuật toán học của mạng Noron có thể xem trong phần phụ lục A.



Hình 2.12: Huấn luyện mạng con RN_1 của hệ thống liên mạng meta-Nơron (m, n, L)

Tiến hành tương tự cho các mạng con còn lại RN_2, RN_3, \dots, RN_m . Đến khi ta thu được các bộ trọng số của các mạng con và output của các mạng con. Kết thúc giai đoạn thứ nhất: huấn luyện cho các mạng con.

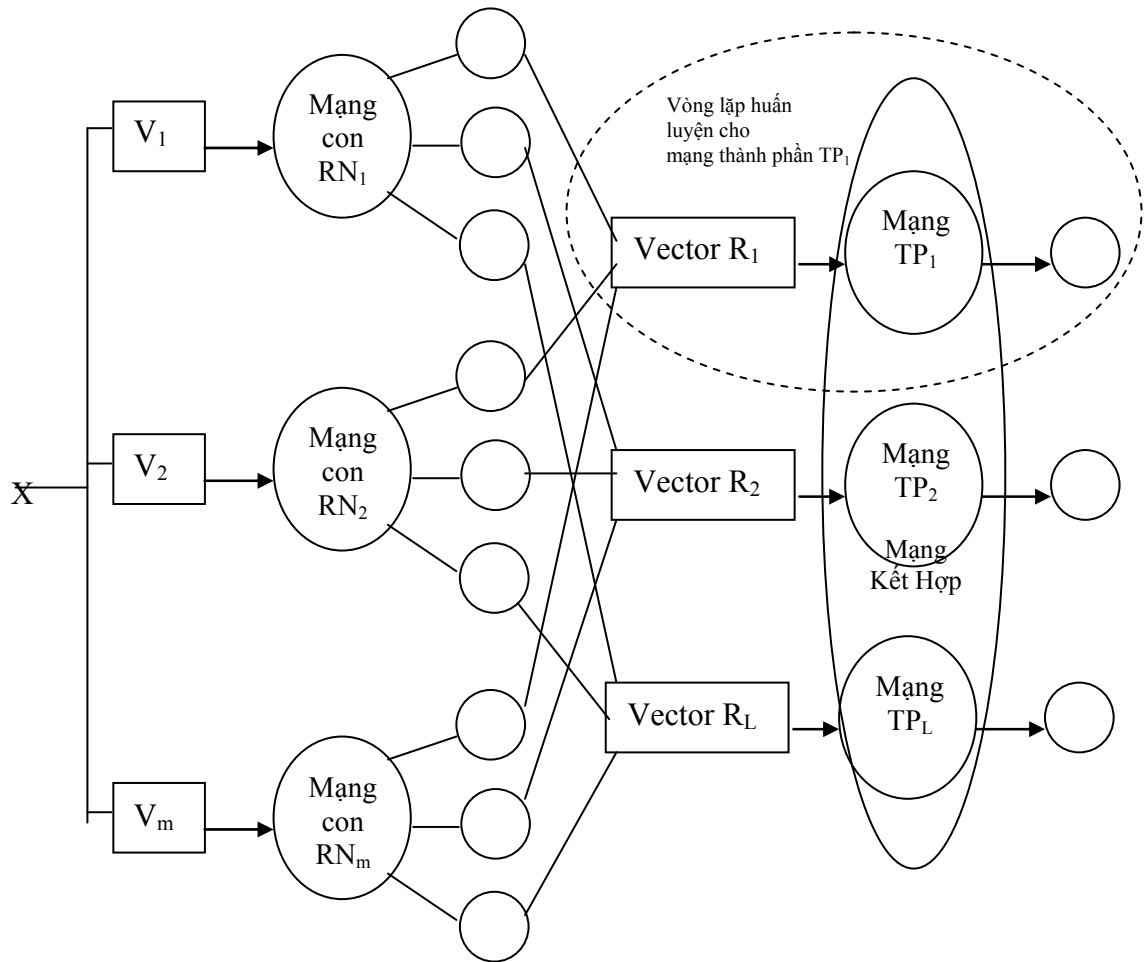
Chẳng hạn, có 822 mẫu mỗi mẫu đặc trưng bởi 4 vector 5 chiều. Trong đó có 201 mẫu (X^1, \dots, X^{201}) thuộc phân hoạch 1 (ví dụ mẫu thuộc về Hạ Long), 367 mẫu (X^{202}, \dots, X^{569}) thuộc về phân hoạch 2 (ví dụ mẫu thuộc về Hà Nội) và 254 mẫu (X^{570}, \dots, X^{822}) thuộc về phân hoạch 3 (ví dụ mẫu thuộc về Nha Trang). Mỗi mẫu X đặc trưng bởi 4 vector nên ta có 4 mạng con. 822 mẫu X , mỗi mẫu lấy vector đặc trưng thứ nhất v_1 , sẽ có 822 vector v_1 . Trong đó có 201 vector v_1 ứng với output kỳ vọng là (1, 0, 0), 367 vector ứng với output kỳ vọng là (0, 1, 0) và 254 vector ứng với kỳ vọng là (0, 0, 1); 822 vector và các output này sẽ dùng để huấn luyện cho mạng con thứ nhất RN_1 .

Hoàn toàn tương tự cho mạng con thứ hai, thứ ba và thứ tư.

2.2.3.1 Huấn luyện cho các mạng toàn cục:

Mạng toàn cục gồm L mạng thành phần TP_i . Quá trình huấn luyện mạng toàn cục: lần lượt huấn luyện cho các mạng thành phần TP_1, TP_2, \dots, TP_L . Từ output kỳ vọng của mẫu X ta dễ dàng xác định output của các mạng thành phần TP_i . Chú ý mạng thành phần TP_i chỉ có 1 nút output duy nhất cũng chính là nút output thứ i qua toàn bộ hệ thống liên mạng. Ví dụ output kỳ vọng của mẫu X là (1, 0, 0...) trường hợp này mẫu X thuộc về phân hoạch Ω_1 thì output kỳ vọng của mạng thành phần TP_1 là 1 còn output của các mạng thành phần còn lại là 0.

Bây giờ ta cần xác định input của mạng thành phần TP_1 để huấn luyện từ mẫu X . Đó chính là các nút output thứ nhất của X qua m mạng con $ro_{11}, ro_{21}, \dots, ro_{m1}$ tạo thành một vector gộp m chiều R_1 . Đưa vector gộp R_1 vào mạng thành phần TP_1 ta sẽ có một giá trị output nào đó với bộ trọng số hiện tại của mạng TP_1 . Sau đó điều chỉnh bộ trọng số của mạng TP_1 để output này gần với output kỳ vọng. Lặp đi lặp lại quá trình này dựa trên tập mẫu X huấn luyện.



Hình 2.13: Huấn luyện mạng thành phần TP_1 của hệ thống liên mạng meta-Nơron (m, n, L)

Sau khi huấn luyện cho mạng thành phần TP_1 của mạng toàn cục ta tiếp tục huấn luyện cho các mạng thành phần còn lại TP_2, \dots, TP_L .

Tiếp tục với ví dụ có 822 mẫu (mỗi mẫu đặc trưng bởi 4 vector, mỗi vector 5 chiều như ví dụ trên phần huấn luyện các mạng con) đã hoàn tất quá trình huấn luyện các mạng con. Nghĩa là đã có bộ trọng số cho 4 mạng con. Xét mẫu X^1 (thuộc về phân hoạch thứ nhất) trong 822 mẫu trên có 4 vector đặc trưng v_1, v_2, v_3, v_4 . Vector đặc trưng thứ nhất v_1 của X^1 qua mạng con thứ nhất sẽ có 3 giá trị ở 3 nút output giả sử là $(h_{l1}=0.7, h_{n1}=0.4, n_{t1}=0.2)$. Tương tự vector v_2, v_3, v_4 qua các mạng con lần lượt cho ra các output giả sử là $(h_{l2}=0.8, h_{n2}=0.3, n_{t2}=0.3)$, $(h_{l3}=0.4, h_{n3}=0.6, n_{t3}=0.1)$, $(h_{l4}=0.5, h_{n4}=0.5, n_{t4}=0.5)$. Với các giá trị giả sử thì này sau quá

trình huấn luyện các mạng con, quyết định đánh giá về mẫu X^1 của mạng con thứ nhất và thừa hai là chính xác còn của mạng con thứ ba và thứ tư thì không chính xác.

Gộp các nút output thứ nhất của tất cả các mạng con thu được vector $R_1=(hl_1, hl_2, hl_3, hl_4)$. Vector R_1 này có giá trị output kỳ vọng là 1, sẽ đưa vào mạng thành phần TP_1 để huấn luyện (chú ý mạng thành phần TP_i chỉ có một nút output duy nhất).

Gộp các nút output thứ hai của tất cả các mạng con thu được vector $R_2=(hn_1, hn_2, hn_3, hn_4)$. Vector R_2 này có giá trị output kỳ vọng là 0, sẽ đưa vào mạng thành phần TP_2 để huấn luyện (chú ý mạng thành phần TP_i chỉ có một nút output duy nhất).

Tiếp tục gộp các nút output thứ ba của tất cả các mạng con thu được vector $R_3=(nt_1, nt_2, nt_3, nt_4)$. Vector R_3 này có giá trị output kỳ vọng là 0, sẽ đưa vào mạng thành phần TP_3 để huấn luyện (chú ý mạng thành phần TP_i chỉ có một nút output duy nhất).

Như vậy có 822 mẫu X thì ta sẽ có 822 vector gộp R_1 đã biết output để đưa vào mạng thành phần TP_1 của mạng toàn cục. Huấn luyện mạng thành phần TP_1 . Tương tự ta cũng có 822 vector R_2, R_3 đã biết output dùng huấn luyện cho mạng thành phần TP_2, TP_3 .

Sau khi huấn luyện xong cho tất cả các mạng thành phần TP_i thì kết thúc giai đoạn huấn luyện thứ hai. Bây giờ có thể chuyển sang giai đoạn đánh giá phân hoạch mẫu X mới.

2.2.4 Quá trình đánh giá một mẫu X qua hệ thống liên mạng

Khi một mẫu X cần phân vào một trong L phân hoạch được đưa vào hệ thống, bộ trích chọn đặc trưng sẽ rút ra m vector đặc trưng v_k với k từ 1 đến m .

Mỗi vector v_k qua một mạng con sẽ cho ta một vector output. V_1 qua mạng Nơron con RN_1 cho ta cho ra các L nút output là $ro_{11}, ro_{12}, \dots, ro_{1L}$. V_2 qua mạng Nơron con RN_2 cho ta cho ra các L nút output là $ro_{21}, ro_{22}, \dots, ro_{2L}$. V_m qua mạng Nơron con RN_m cho ta cho ra các L nút output là $ro_{m1}, ro_{m2}, \dots, ro_{mL}$.

Gộp output thứ nhất của tất cả các mạng con lại thu được vector $R_1=(ro_{11}, ro_{21}, \dots, ro_{m1})$ có m chiều. Vector R_1 qua mạng Nơron thành phần TP_1 của mạng toàn cục cho ra giá trị nút output thứ nhất của toàn hệ thống liên mạng (chính là giá trị nút output duy nhất của mạng thành phần TP_1).

Gộp output thứ hai của tất cả các mạng con lại thu được vector $R_2=(ro_{12}, ro_{22}, \dots, ro_{m2})$ có m chiều. Vector R_2 qua mạng Nơron thành phần TP_2 của mạng toàn cục cho ra giá trị nút output thứ hai của toàn hệ thống liên mạng (chính là giá trị nút output duy nhất của mạng thành phần TP_2).

Cứ tiếp tục gộp đến output thứ L của tất cả các mạng con lại thu được vector $R_L=(ro_{1L}, ro_{2L}, \dots, ro_{mL})$ có m chiều. Vector R_L qua mạng Nơron thành phần TP_L của mạng toàn cục cho ra giá trị nút output thứ L của toàn hệ thống liên mạng (chính là giá trị nút output duy nhất của mạng thành phần TP_L).

Bây giờ ta đã xác định được giá trị của tất cả các nút output của X qua hệ thống liên mạng. Giả sử, mẫu X qua hệ thống liên mạng giá trị của nút output thứ hai là 0.9 có nghĩa là xác suất mẫu X thuộc về phân hoạch thứ $i=2$ là 90%.

2.3 Hệ thống liên mạng meta-Nơron cho bài toán truy vấn ảnh

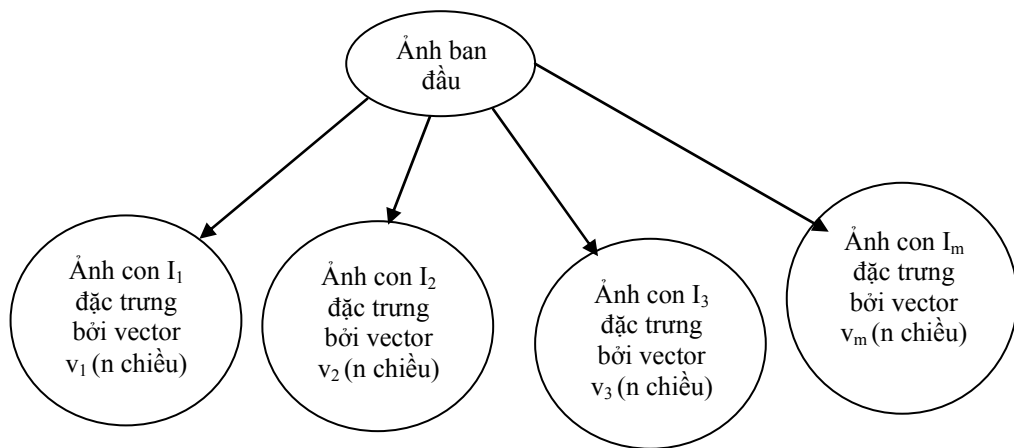
Xét phân hoạch $\Omega = \{\Omega_i \mid 1 \leq i \leq L\}$ phủ kín trên tập ảnh T . Các ảnh giống nhau theo tiêu chí truy vấn cùng thuộc một phân hoạch.

Hàm output của hệ thống mạng liên mạng meta-Nơron sẽ giúp ta xác định mức độ tương đồng giữa hai ảnh như sau: ảnh A qua hệ thống mạng liên mạng

meta-Noron có output là $X=(X_1, X_2,..., X_L)$, ảnh B qua hệ thống mạng liên mạng meta-Noron có output là $Y=(Y_1, Y_2,..., Y_L)$. Độ đo sự tương đồng của A và B:

$$d(A,B) = \min(|X_1-Y_1|, |X_2-Y_2|, \dots, |X_L-Y_L|) \quad (2.5)$$

Mỗi ảnh được tách thành m ảnh con ký hiệu i_1, i_2, \dots, i_m . Mỗi ảnh con i_k (với k từ 1 tới m) sẽ đặc trưng bởi một vector đặc trưng v_k gồm n phần tử $v_k(d_1, d_2, \dots, d_n)$. Như vậy mỗi ảnh gốc ban đầu sẽ được đại diện bởi một bộ gồm m vector đặc trưng mỗi vector có n chiều.



Hình 2.14: Sơ đồ rút trích đặc trưng

Sau đó ta áp dụng một mạng Noron cho từng ảnh con, ký hiệu RN_k cho ảnh con i_k . Mạng Noron RN_k với k từ 1 tới m sẽ có 3 tầng. Trong đó tầng input có n nút, tầng ẩn sẽ có số nút từ 1 đến $2n$ và nút tầng output ký hiệu RO_k . Mạng Noron áp dụng cho các ảnh con chỉ có một tầng ẩn vì đối với các mạng tuyến tính và một số mạng phi tuyến thì có nhiều tầng ẩn thì ta cũng có thể quy về một lớp ẩn. Số nút ở tầng ẩn trong khoảng từ 1 đến $2n$ vì qua thực nghiệm trong một số hệ thống cụ thể thì khi số nút ẩn tăng lên quá $2n$ thì kết quả của quá trình học cũng không thay đổi nhiều.

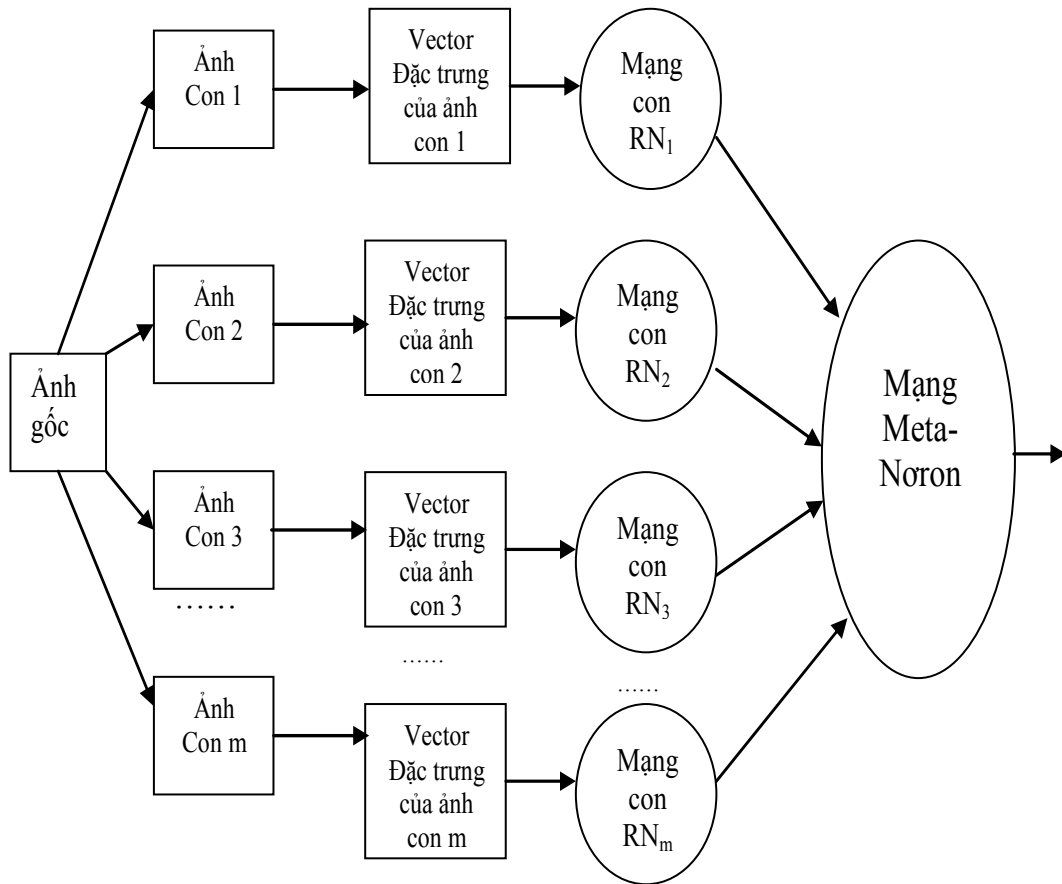
Vấn đề đặt ra là chúng ta sẽ chọn bao nhiêu nút cho tầng ẩn là tối ưu. Trong phạm vi nghiên cứu này chưa tìm ra được công thức toán học để xác định số nút ẩn

tối ưu. Việc xác định số nút ẩn tối ưu cho từng mạng con sẽ được xác định bằng thực nghiệm. Bằng cách tiến hành học nhanh với số bước hữu hạn với số nút ẩn lần lượt từ 1 đến $2n$ và căn cứ vào bảng kết quả tỷ lệ phân lớp chính xác và tỷ lệ bao phủ cao có tổng cao nhất ở số nút ẩn bao nhiêu thì sẽ xây dựng mạng con đó với số nút ẩn tương ứng. Chẳng hạn, giả sử sau khi học nhanh cho ảnh con thứ k và kết quả là với 3 nút ở tầng ẩn thì có tổng tỷ lệ phân lớp chính xác và tỷ lệ bao phủ cao nhất thì ta sẽ xây dựng mạng con RN_k sẽ có 3 nút ở tầng ẩn.

Như vậy ta áp dụng một mạng Noron con ký hiệu RN_k cho từng ảnh con i_k với k từ 1 tới m . Mỗi mạng con này sẽ cho ta một output gọi là RO_k . Ta có m ảnh con, qua m mạng con tương ứng sẽ cho ra m output RO_k .

Sau khi huấn luyện xong tất cả m ảnh con trên toàn cơ sở dữ liệu ảnh. Ta tiến hành tính toán và ghi nhận các output của từng mạng Noron của từng ảnh con. Như vậy bây giờ mỗi ảnh sẽ được đặc trưng bởi vector đặc trưng mới có thành phần là output của các mạng Noron con RN_k cho từng ảnh con với k từ 1 tới m . Và giờ đây ta tiến hành huấn luyện cho mạng Noron toàn cục gọi là mạng meta-Noron với input là các output của các mạng Noron cục bộ cho từng region. Số nút input của mạng meta-Noron sẽ chính là số ảnh con m . Số nút ẩn trong khoảng từ 1 đến $2m+2$ và cũng sẽ được chọn thông qua thực nghiệm như mạng con. Tuy nhiên input của mạng meta-Noron sẽ là các output của các mạng Noron con chứ không phải là trực tiếp đặc trưng của ảnh ban đầu.

Như vậy mạng meta-Noron toàn cục trong tổng thể của hệ thống đóng vai trò quyết định mức độ quan trọng của các mạng Noron cục bộ. Hay nói cách khác là output của các mạng Noron cục bộ góp bao nhiêu phần trăm vào kết quả output cuối cùng sẽ do kết quả huấn luyện của mạng meta-Noron. Điều này rất quan trọng trong bài toán truy vấn ảnh vì tùy vào mỗi loại hình ảnh thì mức độ quan trọng của mỗi ảnh con khác nhau.



Hình 2.15: Kiến trúc hệ thống liên mạng meta-Noron áp dụng trên ảnh

Hệ thống liên mạng meta-Noron là một đóng góp trong suốt hai năm nghiên cứu luận văn này, dựa trên cơ sở cải tiến phương pháp kết hợp thuật giải di truyền với mạng Noron áp dụng cho bài toán nhận dạng vân tay và ký tự đã được trình bày trong luận án Tiến sĩ của tác giả Lê Hoàng Thái. Tác giả đã sử dụng thuật giải di truyền để đánh giá mức độ tin cậy của các mạng Noron con. Trong phạm vi luận văn này sử dụng một mạng Noron toàn cục gọi là meta-Noron thay cho việc sử dụng thuật giải di truyền để xác định hệ số tin cậy của các mạng con. Bởi vì trong bài toán truy vấn ảnh tổng quát, tùy thuộc vào từng loại ảnh cũng như từng mục đích truy vấn cụ thể mà mức độ quan trọng của các ảnh con sẽ khác nhau. Thậm chí cùng một tập dữ liệu ảnh nhưng mục đích truy vấn khác nhau thì mức độ quan trọng của các ảnh con cũng có thể khác nhau. Do đó hệ số tin cậy của các kết quả do

mạng con tạo ra phải được xác định qua quá trình “học” thì mới tốt cho bài toán truy vấn ảnh tổng quát.

Tóm lại, để giải bài toán truy vấn ảnh, luận văn đề xuất sử dụng phương pháp kết hợp kết quả các mạng Noron con lại với nhau. Tuy nhiên luận văn không dùng thuật giải di truyền hay là logic mờ như các tác giả trước đây mà sử dụng một mạng Noron gọi là meta-Noron để kết hợp các mạng Noron con lại với nhau. Mạng meta-Noron sẽ kết hợp các kết quả output của các mạng con lại thành đầu vào input cho mạng meta-Noron. Từ đó qua mạng meta-Noron toàn cục ta thu được kết quả cuối cùng cho ra bởi toàn bộ hệ thống liên mạng meta-Noron.

Phương pháp kết hợp bằng hệ thống liên mạng meta-Noron thực hiện sự kết hợp output của các mạng con bằng các mạng thành phần của mạng meta-Noron toàn cục. Cụ thể các output thứ i của tất cả các mạng con sẽ kết hợp lại thành vector input của mạng thành phần TP_i của mạng meta-Noron toàn cục. Mạng thành phần TP_i sẽ xác định giá trị nút output thứ i của toàn hệ thống liên mạng meta-Noron. Ngoài ra, để phát triển ứng dụng hệ thống liên mạng chỉ cần người lập trình phát triển biết kiến thức thuần túy mạng Noron.

Hơn thế nữa, nếu việc rút trích các đặc trưng cơ bản tốt sau khi vào mạng Noron thành phần có thể rút trích ra được các ngữ nghĩa trong ảnh như có thuyền, có hoa... thì có thể triển khai thành hệ thống truy vấn ảnh dựa trên ngữ nghĩa ảnh.

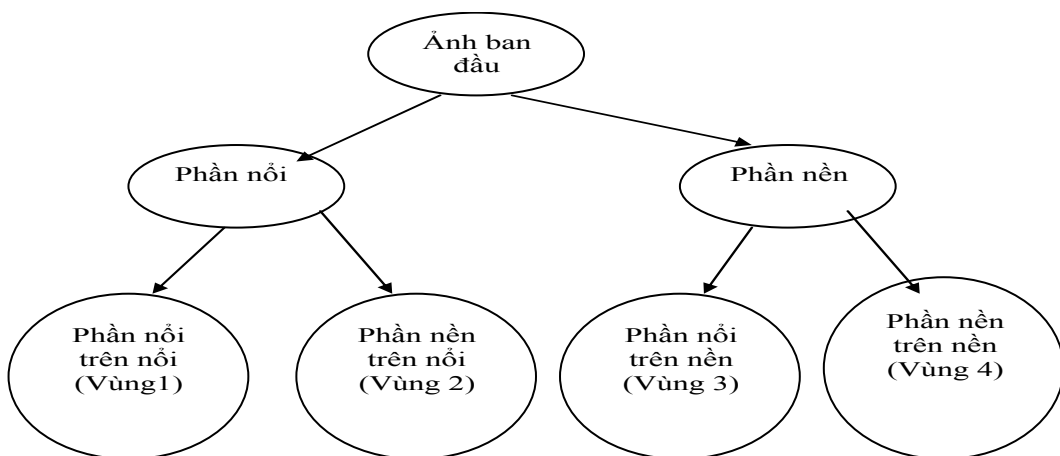
Chương 3: MỘT ÁP DỤNG CỦA HỆ THỐNG LIÊN MẠNG META-NƠRON

Chương này trình bày một áp dụng của hệ thống liên mạng meta-Nơron đã đề xuất ở chương 2 vào bài toán truy vấn ảnh phong cảnh vùng du lịch. Tập ảnh huấn luyện bao gồm 822 ảnh với 201 ảnh phong cảnh Hạ Long tìm trên mạng, 367 ảnh phong cảnh Hà Nội chụp bằng máy ảnh kỹ thuật số trong lần ra Hà Nội vào tháng 8/2007, và 254 ảnh phong cảnh Nha Trang cũng chụp bằng máy ảnh kỹ thuật số trong nhiều lần về quê Nha Trang.

3.1 Rút trích các đặc trưng của ảnh

Vì đầu vào của mạng Nơron là dữ liệu lượng tử nên trước tiên từ tập tin hình ảnh ta cần phải rút ra vector đặc trưng chứa các đặc trưng trong ảnh. Từ tập cơ sở dữ liệu ảnh ta sẽ rút ra tập vector đặc trưng để đưa vào huấn luyện.

Cụ thể mỗi bức ảnh sẽ phân tách thành 4 ảnh con như sau: đầu tiên chia ảnh thành hai phần là phần nền và phần nổi dựa trên mức xám. Sau đó, phần nổi lại chia làm hai phần nền và nổi trên phần nổi. Đồng thời phần nền cũng tiếp tục chia thành hai phần nền và nổi của nền.



Hình 3.1: Sơ đồ rút trích đặc trưng

Phần nổi sẽ bao gồm những pixel của ảnh gốc có mức xám cao hơn mức xám trung bình. Phần nền bao gồm những pixel có mức thấp hơn mức xám trung bình trong ảnh gốc ban đầu. Phần 1 (phần nổi của phần nổi) sẽ bao gồm những pixel có mức xám cao hơn trung bình trong phần nổi của ảnh gốc. Phần 2 (phần nền của phần nổi) sẽ bao gồm những pixel có mức xám thấp hơn trung bình trong phần nổi của ảnh gốc. Phần 3 (phần nổi của phần nền) sẽ bao gồm những pixel có mức xám cao hơn trung bình trong phần nền của ảnh gốc. Phần 4 (phần nền của phần nền) sẽ bao gồm những pixel có mức xám thấp hơn trung bình trong phần nền của ảnh gốc.

Ứng với mỗi ảnh con sẽ rút ra vị trí của trọng tâm, tỷ lệ của diện tích ảnh con so với ảnh gốc ban đầu và chuyển từ không gian màu RGB sang không gian màu LHC.

Tại mỗi vùng rút ra các giá trị sau: vị trí của trọng tâm (1,2,3,4 tương ứng với góc trên bên trái, góc trên bên phải, góc dưới bên trái, góc dưới bên phải)

1	2
3	4

Và tỷ lệ giữa diện tích của vùng so với toàn ảnh = tổng số pixel của vùng chia cho tổng số pixel của ảnh. Đồng thời lấy ra màu L H C của tâm theo công thức chuyển đổi sau:

$$\begin{aligned}
 L &= 116 \left(\frac{Y}{Y_0} \right)^{1/3} - 16 \\
 H &= \tan^{-1} \left(\frac{b^*}{a^*} \right) \\
 C &= \sqrt{(a^*)^2 + (b^*)^2} \\
 a^* &= 500 \left\{ \left(\frac{X}{X_0} \right)^{1/3} - \left(\frac{Y}{Y_0} \right)^{1/3} \right\} \\
 b^* &= 200 \left\{ \left(\frac{Y}{Y_0} \right)^{1/3} - \left(\frac{Z}{Z_0} \right)^{1/3} \right\} \\
 \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} &= \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5943 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}
 \end{aligned} \tag{3.1}$$

Vậy mỗi ảnh gốc ban đầu sẽ được đặc trưng bởi 4 vector, mỗi vector có 5 phần tử đặc trưng cho ảnh con tương ứng. Hay nói cách khác ma trận đặc trưng của mỗi ảnh gốc ban đầu là ma trận 4×5 (4 hàng 5 cột).

Hàng đầu tiên gồm 5 ô tương ứng với 5 đặc trưng của ảnh con thứ nhất (phần nổi của phần nổi), hàng thứ hai gồm 5 ô tương ứng với 5 đặc trưng của ảnh con thứ hai (phần nền của phần nổi), hàng thứ ba gồm 5 ô tương ứng với 5 đặc trưng của ảnh con thứ ba (phần nổi của phần nền), hàng thứ tư gồm 5 ô tương ứng với 5 đặc trưng của ảnh con thứ tư (phần nền của phần nền).

Cột thứ nhất ứng với vị trí của trọng tâm, cột thứ hai chứa tỷ lệ diện tích ảnh con so với ảnh gốc ban đầu, cột thứ ba là màu L, cột thứ tư là màu H, và cột thứ năm là màu C của trọng tâm.



Hình 3.2: Ảnh đầu vào

Bảng 3.1: Biểu diễn ảnh thành các đặc trưng

	Vị trí	Tỷ lệ diện tích	Màu L	Màu H	Màu C
Ảnh con 1	3	0.2622148	159.0667	0.59617	4.534818
Ảnh con 2	1	0.33285925	226.805	0.41962	0.92006
Ảnh con 3	1	0.32861853	55.96102	0.65679	9.151924
Ảnh con 4	1	0.07630741	116.4482	0.74906	6.609669

Như vậy, mỗi ảnh sẽ được đặc trưng bởi 4 vector đặc trưng, mỗi vector 5 chiều. Các vector đặc trưng này sẽ là đầu vào cho hệ thống liên mạng meta-Noron với bộ tham số ($m=4$, $n=5$, $L=3$) sẽ cho ra 3 giá trị kết xuất. Các giá trị kết xuất này giúp ta xác định sự tương đồng giữa hai ảnh dùng hàm độ đo sự tương đồng theo công thức (2.5). Giá trị của hàm độ đo sự tương đồng giúp hệ thống xác định kết quả truy vấn: ảnh trả ra là ảnh giống với ảnh truy vấn nhất trong cơ sở dữ liệu ảnh dựa theo giá trị độ đo sự tương đồng trên (hai ảnh giống nhau thì giá trị hàm độ đo sự tương đồng càng nhỏ).

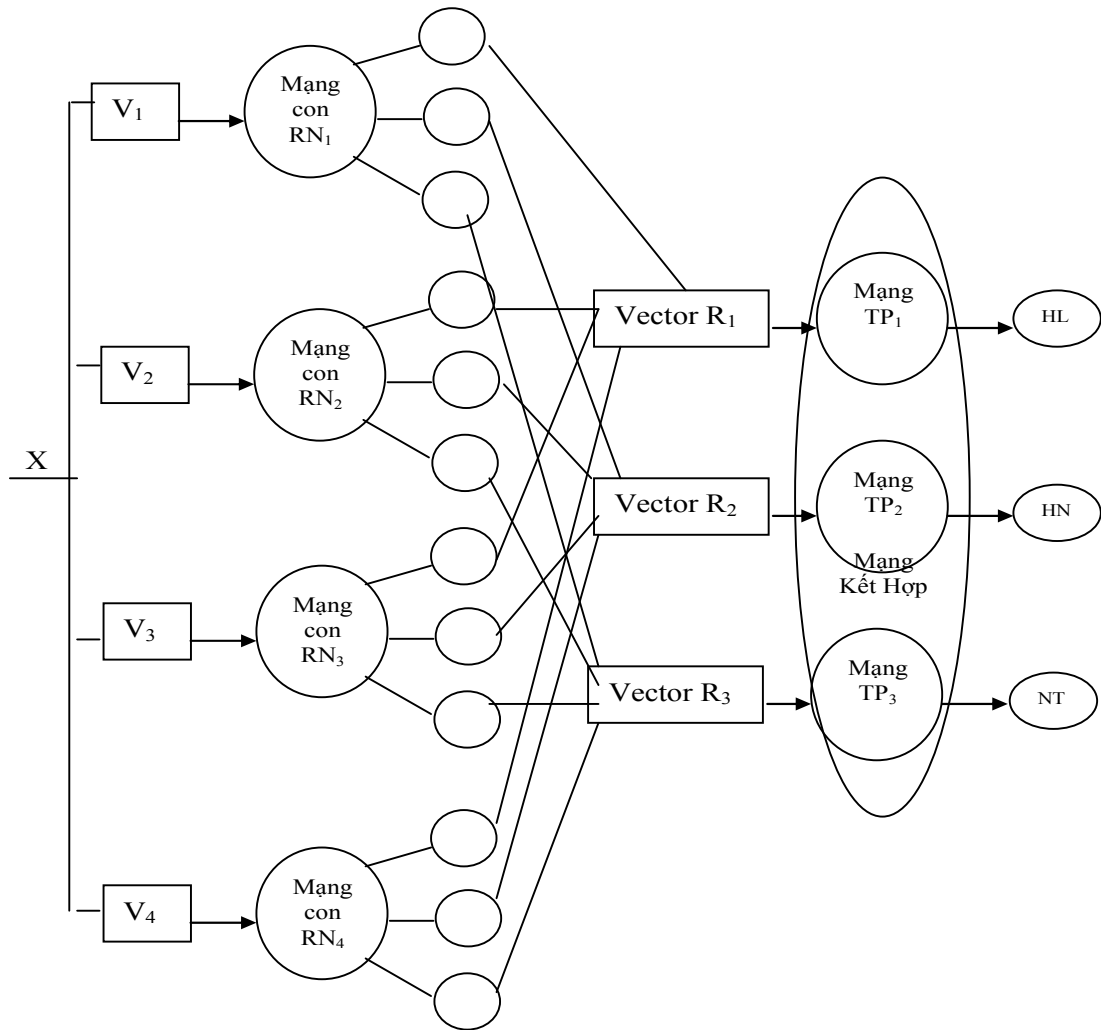
Trong phạm vi luận văn này tập trung vào việc xây dựng hệ thống liên mạng meta-Noron áp dụng cho bài toán truy vấn ảnh tổng quát, không tập trung vào việc cải tiến các thuật toán rút trích đặc trưng từ ảnh cho tối ưu. Nếu có điều kiện kết hợp với các công trình nghiên cứu khác về rút trích đặc trưng hay các đối tượng trong ảnh thì kết quả sẽ cao hơn. Tuy nhiên, phần lớn các công trình rút ra các đối tượng trong ảnh có kết quả tốt thì phải xét trong phạm vi một bài toán cụ thể, một loại hình ảnh cụ thể. Ví dụ: rút trích ra các đặc trưng của khuôn mặt hay rút trích ra gương mặt người trong ảnh chân dung...

3.2 Kiến trúc hệ thống liên mạng meta-Noron

Mỗi ảnh được tách thành 4 ảnh con, mỗi ảnh con rút một vector đặc trưng 5 chiều. Vậy mẫu X trong trường hợp này được đặc trưng bởi $m=4$ vector, mỗi vector $n=5$ chiều. Phân hoạch X vào một trong $L=3$ lớp Hạ Long, Hà Nội hay Nha Trang.

Các thành phần trong hệ thống liên mạng gồm: $m=4$ mạng Noron con RN_1 , RN_2 , RN_3 , RN_4 và một mạng toàn cục gồm $L=3$ mạng thành phần TP_1 , TP_2 , TP_3 . Trong đó các mạng Noron con cũng như các mạng thành phần của mạng toàn cục đều dùng mạng lan truyền ngược 3 lớp (1 lớp input, 1 lớp ẩn và 1 lớp xuất) với hàm truyền Sigmoid. Chi tiết về mạng lan truyền ngược với hàm Sigmoid được trình bày trong phụ lục A.

Cụ thể các mạng con RN_k với $k=1,2,3,4$ có $n=5$ nút input và $L=3$ nút output. Số Neuron nút ẩn xác định bằng quá trình thực nghiệm. Các mạng thành phần TP_i với $i=1,2,3$ của mạng toàn cục sẽ có $m=4$ nút input và 1 nút output duy nhất.



Hình 3.3: Hệ thống liên mạng ($m=4, n=5, L=3$)

3.3 Xây dựng hệ thống liên mạng meta-Neuron

Output của hệ thống liên mạng meta-Neuron sẽ giúp xác định mức độ tương đồng giữa hai ảnh: ảnh A qua hệ thống liên mạng meta-Neuron có output là X, ảnh B qua mạng Neuron có output là Y. Độ đo sự tương đồng của A và B tính theo công thức độ đo sự tương đồng (2.5) đã trình bày ở chương 2.

Từ cơ sở dữ liệu 822 ảnh phong cảnh của 3 vùng Hạ Long, Hà Nội và Nha Trang. Tiến hành phân tách mỗi ảnh thành 4 ảnh con như đã nói ở phần trên. Mỗi ảnh con rút ra vector đặc trưng của nó. Ghi nhận các đặc trưng vào cơ sở dữ liệu.

Tiến hành quá trình huấn luyện cho mạng con thứ nhất RN_1 , để phân lớp ảnh dựa vào các đặc trưng của ảnh con thứ nhất trong các tập ảnh huấn luyện. Tiến hành học nhanh với số bước hữu hạn 20000 bước, cho mạng con thứ nhất RN_1 để phân lớp ảnh thuộc về Hạ Long hay không với số nút ẩn từ 1-10. Do nhược điểm của hệ thống liên mạng gồm nhiều mạng Noron ($m+L$). Mỗi mạng Noron cần thời gian huấn luyện tương đối dài, số nút ẩn tối ưu cho mỗi mạng chưa có công thức xác định trước. Vì vậy để nhanh chóng xác định số nút ẩn tốt cho các mạng Noron, luận văn tiến hành thực hiện quá trình học nhanh với số bước hữu hạn để chọn ra số nút ẩn tối ưu cho các mạng Noron trong hệ thống liên mạng.

Kết quả quá trình học nhanh này như sau:

Bảng 3.2: Thực nghiệm phân hoạch Hạ Long theo mạng con

HALONG	Chọn 8 3 6 3									
Region1										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.0%	28.3%	27.4%	25.8%	25.7%	26.1%	26.6%	25.9%	28.3%	28.6%
Recall	0.0%	68.7%	78.6%	74.1%	75.1%	74.6%	74.6%	80.6%	69.7%	70.6%
Region2										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.0%	23.7%	23.3%	24.4%	24.7%	25.1%	25.2%	24.0%	26.5%	24.6%
Recall	0.0%	84.6%	90.5%	77.1%	69.7%	70.1%	68.7%	88.1%	57.2%	68.7%
Region3										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.0%	28.6%	24.6%	29.6%	27.9%	29.1%	29.7%	24.7%	29.5%	25.9%
Recall	0.0%	53.7%	82.1%	66.2%	77.1%	79.1%	78.1%	81.6%	64.7%	75.1%
Region4										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.0%	25.0%	25.9%	24.9%	25.4%	25.4%	25.5%	25.5%	25.5%	25.3%
Recall	0.0%	73.1%	90.5%	80.6%	83.1%	82.6%	82.6%	88.6%	76.6%	81.1%

Tiến hành tương tự cho phân lớp ảnh Hà Nội thu được bảng kết quả thực nghiệm như sau:

Bảng 3.3: Thực nghiệm phân hoạch Hà Nội theo mạng con

HANOI	Chọn 4 2 7 1									
Region1										
Node	1	2	3	4	5	6	7	8	9	10
Precision	44.65%	45.06%	45.11%	44.81%	44.81%	44.81%	44.84%	44.76%	44.81%	44.65%
Recall	100.00%	99.46%	99.18%	100.00%	100.00%	100.00%	99.46%	100.00%	100.00%	100.00%
Region2										
Node	1	2	3	4	5	6	7	8	9	10
Precision	44.65%	45.07%	44.76%	44.76%	44.74%	44.74%	44.83%	44.65%	44.74%	44.65%
Recall	100.00%	99.73%	100.00%	100.00%	99.73%	99.73%	99.18%	100.00%	99.73%	100.00%
Region3										
Node	1	2	3	4	5	6	7	8	9	10
Precision	44.65%	47.36%	49.27%	55.13%	48.09%	49.29%	53.29%	49.57%	47.68%	49.10%
Recall	100.00%	90.46%	91.83%	90.74%	92.64%	94.55%	92.64%	94.55%	95.10%	96.19%
Region4										
Node	1	2	3	4	5	6	7	8	9	10
Precision	44.65%	44.82%	45.71%	45.83%	45.83%	46.46%	47.00%	45.73%	45.72%	44.61%
Recall	100.00%	89.65%	94.28%	95.91%	94.28%	96.46%	94.01%	96.19%	94.55%	98.09%

Tương tự được các bảng kết quả thực nghiệm cho quá trình học nhanh dùng phân lớp ảnh Nha Trang:

Bảng 3.4: Thực nghiệm phân hoạch Nha Trang theo mạng con

NHATRANG		Chọn 9 2 9 10								
Region1										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.00%	68.71%	78.03%	76.54%	60.51%	86.05%	78.51%	64.13%	36.72%	36.67%
Recall	0.00%	44.09%	40.55%	24.41%	65.75%	29.13%	37.40%	46.46%	99.61%	99.61%
Region2										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.00%	35.08%	42.46%	35.12%	31.36%	39.64%	43.51%	35.29%	30.89%	31.18%
Recall	0.00%	97.64%	89.76%	46.46%	35.43%	86.61%	79.13%	96.85%	99.61%	98.82%
Region3										
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.00%	30.77%	33.33%	37.50%	22.22%	33.33%	41.67%	30.23%	34.71%	38.85%
Recall	0.00%	1.57%	1.57%	2.36%	12.60%	0.39%	1.97%	5.12%	95.67%	90.55%
Region4	0.0000	0.3234	0.3491	0.3986	0.3482	0.3373	0.4364	0.3535	1.3038	1.2940
Node	1	2	3	4	5	6	7	8	9	10
Precision	0.00%	14.55%	31.03%	33.33%	33.89%	45.45%	41.51%	32.32%	34.75%	35.54%
Recall	0.00%	15.35%	7.09%	7.09%	55.51%	1.97%	8.66%	50.00%	100.00%	99.21%

Tổng hợp các số liệu trên, có thể nhận thấy giữa tỷ lệ chính xác và tỷ lệ bao phủ tồn tại mối quan hệ tỷ lệ nghịch. Muốn tỷ lệ chính xác cao thì tỷ lệ bao phủ sẽ thấp và ngược lại muốn tỷ lệ bao phủ cao thì tỷ lệ chính xác sẽ thấp. Vì mục đích áp dụng cho bài toán truy vấn ảnh sau này, nên trong quá trình khởi tạo và huấn luyện mạng sẽ hướng đến việc nâng cao tỷ lệ bao phủ, chấp nhận tỷ lệ chính xác thấp. Bởi lý do khi đưa ra một câu truy vấn tìm kiếm, cho dù là cho thông tin thị giác ảnh hay thông tin thông thường, thì người dùng luôn muốn kết quả cần tìm phải xuất hiện trong các kết quả trả về.

Ví dụ: có ảnh A gần giống với các ảnh A_1, A_2, A_3, A_4, A_5 trong cơ sở dữ liệu ảnh cần tìm và không giống với các ảnh B_1, B_2, B_3, B_4, B_5 . Khi tỷ lệ bao phủ cao và tỷ lệ chính xác thấp thì tập kết quả trả ra sẽ có thể là 5 ảnh giống và cả 5 ảnh khác. Vậy tỷ lệ bao phủ sẽ là 5 ảnh chính xác trả ra chia cho tổng số ảnh chính xác trong cơ sở dữ liệu bằng 100%. Trong khi đó tỷ lệ chính xác chỉ bằng 50% bằng 5 ảnh chính xác trả chia cho tổng số ảnh trả ra là 10. Tuy nhiên với tỷ lệ bao phủ cao như thế thì người dùng thông qua thao tác tìm kiếm và loại bỏ thông tin sai sẽ chắc chắn tìm ra ảnh cần thiết trong cơ sở dữ liệu.

Ngược lại, nếu nâng tỷ lệ chính xác lên 100% và chấp nhận giảm tỷ lệ bao phủ thì kết quả trả ra những ảnh giống ảnh A chỉ có A_1, A_2, A_3 và không có thêm ảnh B nào thì tỷ lệ chính xác sẽ là 100% và tỷ lệ bao phủ còn 60%. Như thế nếu ảnh cần tìm của người dùng là ảnh A_4 chẳng hạn thì không thể sử dụng kết quả tìm kiếm của hệ thống để lấy ra thông tin cần thiết.

Chính vì lý do ứng dụng trên mà khi xây dựng, huấn luyện và thiết lập các mạng Noron phân lớp theo hướng tỷ lệ bao phủ cao và chấp nhận tỷ lệ chính xác ở mức thấp bằng cách thiết lập các trọng số ban đầu với tỷ lệ bao phủ cao tối đa.

Sau khi có cấu hình mạng tối ưu (tìm ra số nút ẩn tốt) cho tất cả các mạng con. Ta tiến hành huấn luyện và ghi nhận bộ trọng số của các mạng con. Trên cơ sở đó, tiến hành tính toán và ghi nhận các output của từng mạng Noron con cho toàn

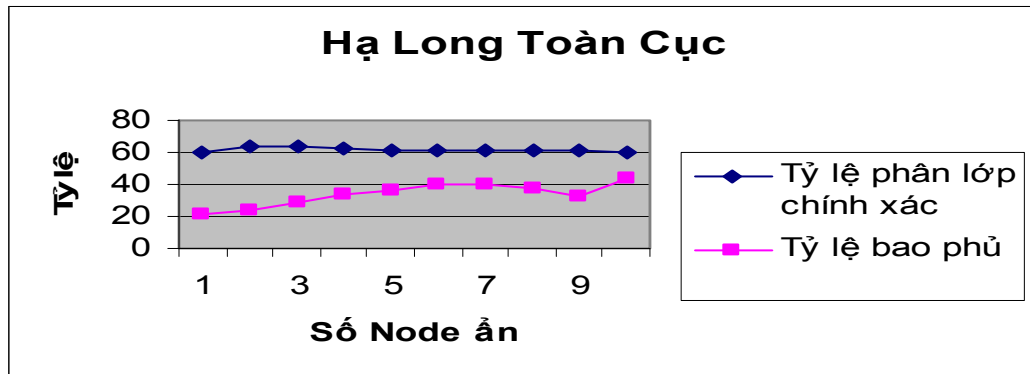
cơ sở dữ liệu ảnh. Các output của các mạng con sẽ là dữ liệu để huấn luyện cho mạng meta-Noron toàn cục. Như vậy lúc này mỗi ảnh được đặc trưng bởi vector đặc trưng mới có thành phần là output của các mạng Noron cho từng ảnh con. Và giờ đây ta tiến hành huấn luyện cho mạng Noron toàn cục meta-Noron với input là các output của các mạng Noron cục bộ cho từng ảnh con. Quá trình huấn luyện của mạng meta-Noron cũng tương tự như các mạng con: học nhanh tìm số nút ẩn tối ưu, học với số nút ẩn tối ưu, lưu giữ bộ trọng số và cập nhật vào cơ sở dữ liệu. Lưu ý quá trình huấn luyện cho mạng meta-Noron toàn cục sẽ không thay đổi bộ trọng số đã học tối ưu của các mạng con.

Tiến hành học nhanh cho mạng meta-Noron toàn cục với số nút ẩn từ 1- 10 $= 2 \times 4 + 2$ để tìm số nút ẩn tối ưu cho mạng meta-Noron toàn cục. Ta có bảng kết quả tỷ lệ phân lớp chính xác và tỷ lệ bao phủ như sau:

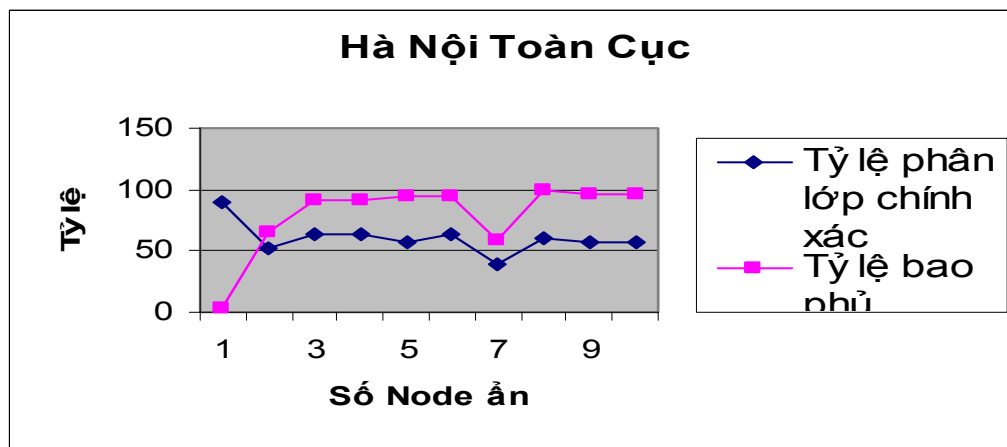
Bảng 3.5: Thực nghiệm phân hoạch qua toàn hệ thống liên mạng

Số nút ẩn	1	2	3	4	5	6	7	8	9	10
HL										
Precision	60.56	63.16	63.33	62.73	61.34	61.54	61.36	61.29	61.68	60
HL Recall	21.39	23.88	28.36	34.33	36.32	39.8	40.3	37.81	32.84	43.28
10	81.95	87.04	91.69	97.06	97.66	101.34	101.66	99.1	94.52	103.28
HN										
Precision	90	52.75	63.88	63.93	57.71	63.21	38.75	59.87	57.24	57.42
HN Recall	2.45	65.4	91.55	91.28	94.82	94.55	59.13	99.18	95.91	95.91
8	92.45	118.15	155.43	155.21	152.53	157.76	97.88	159.05	153.15	153.33
NT										
Precision	42.78	42.44	41.84	42.18	36.71	41.6	41.71	42.14	41.69	41.67
NT Recall	64.17	62.99	62.6	62.6	45.67	65.35	65.35	66.54	56.3	64.96
8	106.95	105.43	104.44	104.78	82.38	106.95	107.06	108.68	97.99	106.63

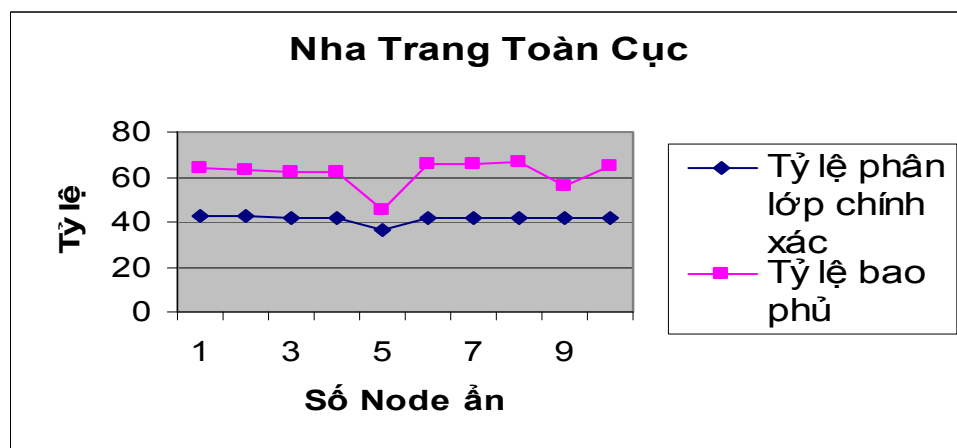
Với bảng kết quả nêu trên, các biểu đồ kết quả sau quá trình học nhanh thể hiện như sau:



Hình 3.4: Biểu đồ thực nghiệm mạng phân lớp Hạ Long



Hình 3.5: Biểu đồ thực nghiệm mạng phân lớp Hà Nội



Hình 3.6: Biểu đồ thực nghiệm mạng phân lớp Nha Trang

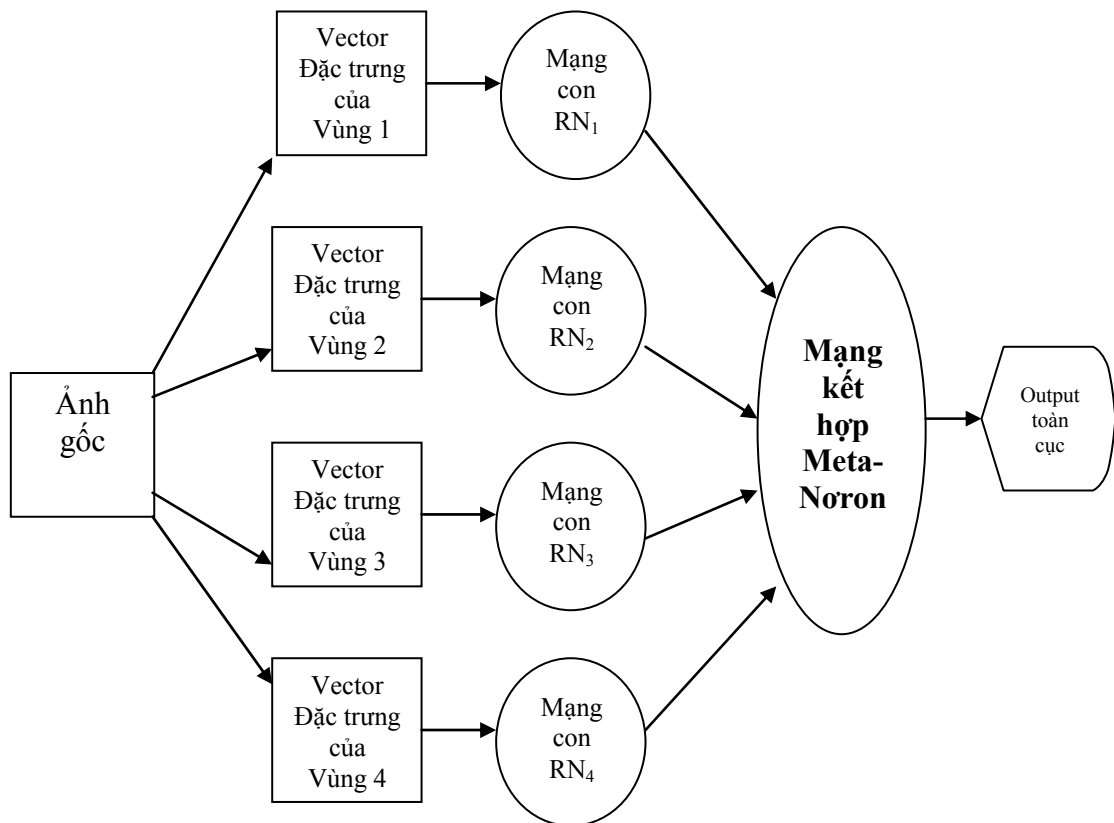
Từ kết quả trên sẽ chọn ra số nút ẩn tối ưu cho các mạng thành phần của mạng meta-Noron toàn cục TP_1 , TP_2 , TP_3 lần lượt là 10,8,8 vì cho kết quả tỷ lệ phân lớp chính xác và bao phủ tốt trong trường hợp này. Qua biểu đồ trên có thể nhận thấy ngay là kết quả của toàn hệ thống sau khi qua mạng meta-Noron cho kết quả cao hơn dù chỉ qua quá trình học nhanh.

Sau đó huấn luyện mạng meta-Noron toàn cục với input chính là output của các mạng con cho 822 ảnh trong cơ sở dữ liệu một cách đầy đủ nghĩa là huấn luyện cho đến khi hệ số lỗi chấp nhận được tức nhỏ hơn 5% hay vượt quá số bước giới hạn. Sau khi huấn luyện xong thì lưu lại các trọng số của mạng meta-Noron toàn cục trong các file: tcHLTotal.bpn, tcHNTotal.bpn, tcNTTotal.bpn.

Sau đó tiến hành cập nhật các output qua mạng meta-Noron toàn cục của 822 ảnh huấn luyện vào cơ sở dữ liệu làm thông tin để sau này tiến hành truy vấn.

Lúc này, một ảnh truy vấn bất kỳ dùng qua các mạng con sẽ cho ra các output, rồi các output này qua mạng meta-Noron toàn cục sẽ cho ra các kết quả cuối cùng. Kết quả cuối cùng của ảnh truy vấn sẽ so sánh với kết quả cuối cùng của các ảnh huấn luyện đã có trong cơ sở dữ liệu. Nếu hai kết quả này gần nhau theo công thức độ đo sự tương đồng (2.5) thì hai ảnh sẽ xem là giống nhau. Từ đó suy ra vùng du lịch của ảnh truy vấn.

Như vậy mạng meta-Noron toàn cục trong tổng thể của hệ thống đóng vai trò quyết định mức độ quan trọng của các mạng Noron cục bộ. Hay nói cách khác là output của các mạng Noron cục bộ góp bao nhiêu phần trăm vào kết quả output cuối cùng sẽ do kết quả huấn luyện của mạng meta-Noron toàn cục. Điều này rất quan trọng trong bài toán truy vấn ảnh vì tùy vào mỗi loại hình ảnh thì mức độ quan trọng của mỗi region khác nhau.



Hình 3.7: Tổng quan quá trình xử lý của hệ thống liên mạng cụ thể ($m=4$, $n=5$, $L=3$)

Trong đó vector đặc trưng của các ảnh con gồm 5 thành phần là: vị trí tâm của ảnh con, tỷ lệ diện tích ảnh con, màu L, màu H và màu C theo công thức đã trình bày trong phần rút trích đặc trưng mục 3.1. Như vậy, mỗi ảnh qua hệ thống liên mạng meta-Nơron sẽ cho kết quả output cuối cùng. Hai ảnh được xem là giống nhau nếu có kết quả output cuối cùng qua hệ thống liên mạng meta-Nơron có giá trị gần nhau theo công thức độ đo sự tương đồng (2.5) trong chương 2.

Chương 4: CHƯƠNG TRÌNH ỨNG DỤNG

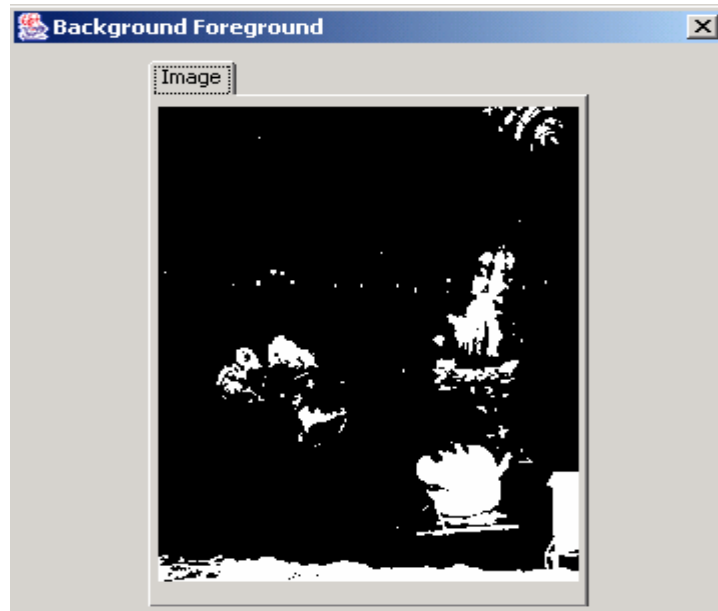
4.1 Quá trình xử lý ảnh trước khi đưa vào hệ thống liên mạng

Từ ảnh gốc ban đầu sẽ tách ra phần nền và phần nổi. Sau đó lại tách phần nền làm hai phần và phần nổi làm hai phần giữa vào mức xám trên cục bộ. Chạy chương trình ImageGUI.class chọn: File→Open



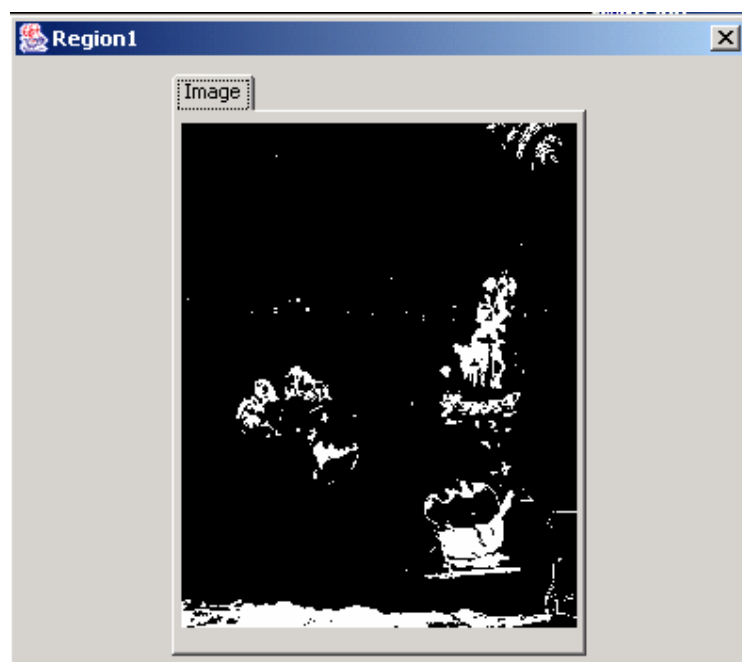
Hình 4.1: Ảnh gốc dùng truy vấn

Để xem kết quả việc tách ảnh gốc thành phần nền và phần nổi, vào Tool → Background ForeGround



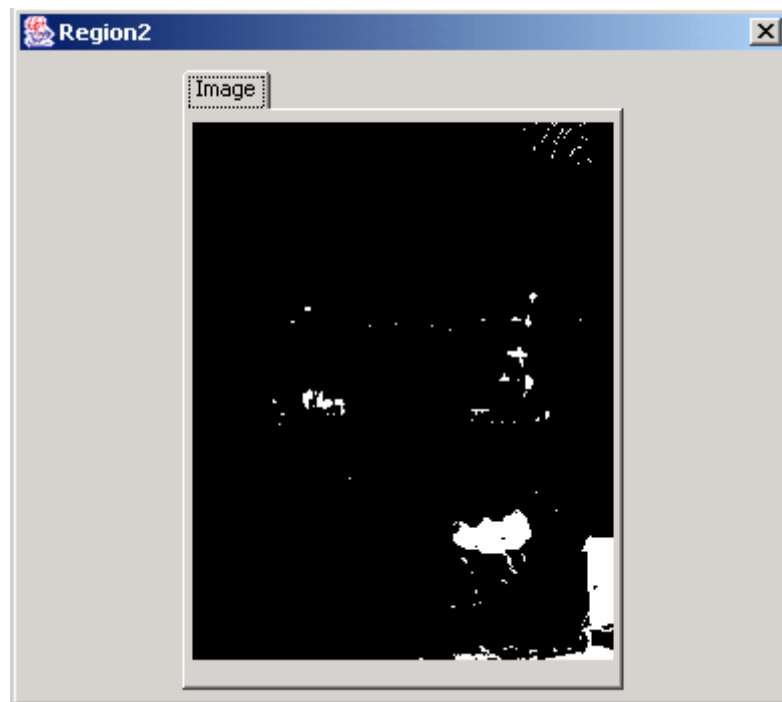
Hình 4.2: Phần nền và phần nổi

Để xem kết quả tách ra ảnh con thứ nhất từ ảnh gốc, vào Tool → Region1.



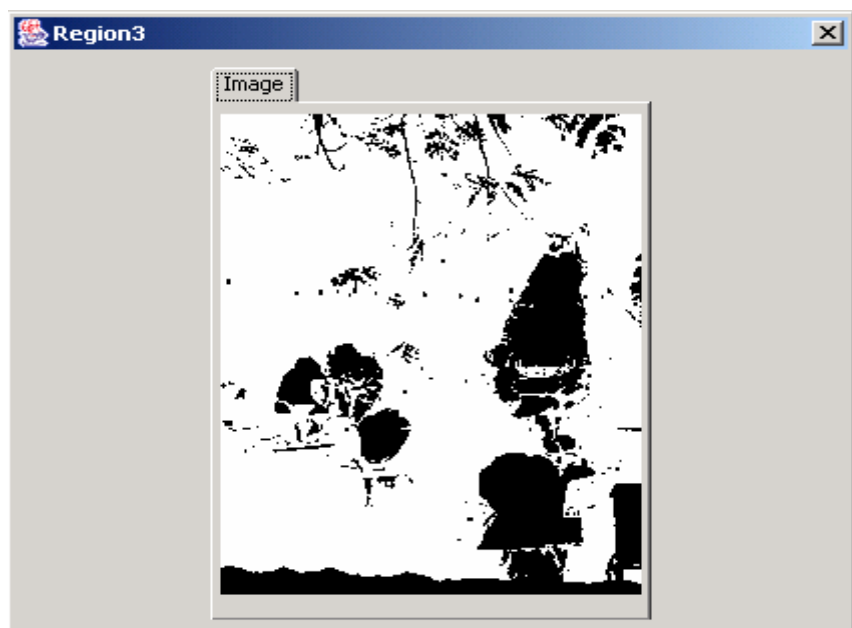
Hình 4.3: Vùng 1

Để xem kết quả tách ra ảnh con thứ hai từ ảnh gốc, vào Tool→Region 2



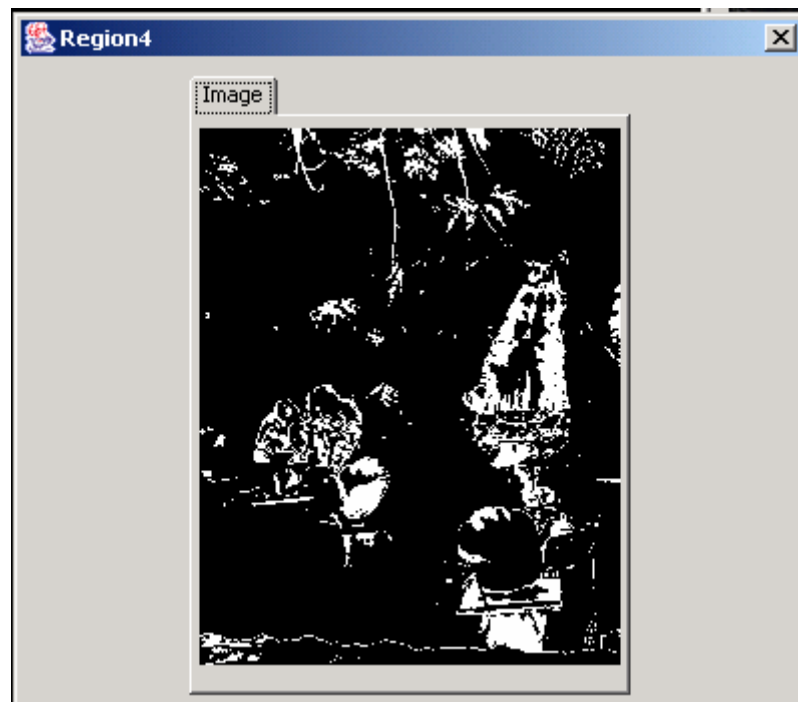
Hình 4.4: Vùng 2

Để xem kết quả tách ra ảnh con thứ ba từ ảnh gốc, vào Tool→Region 3



Hình 4.5: Vùng 3

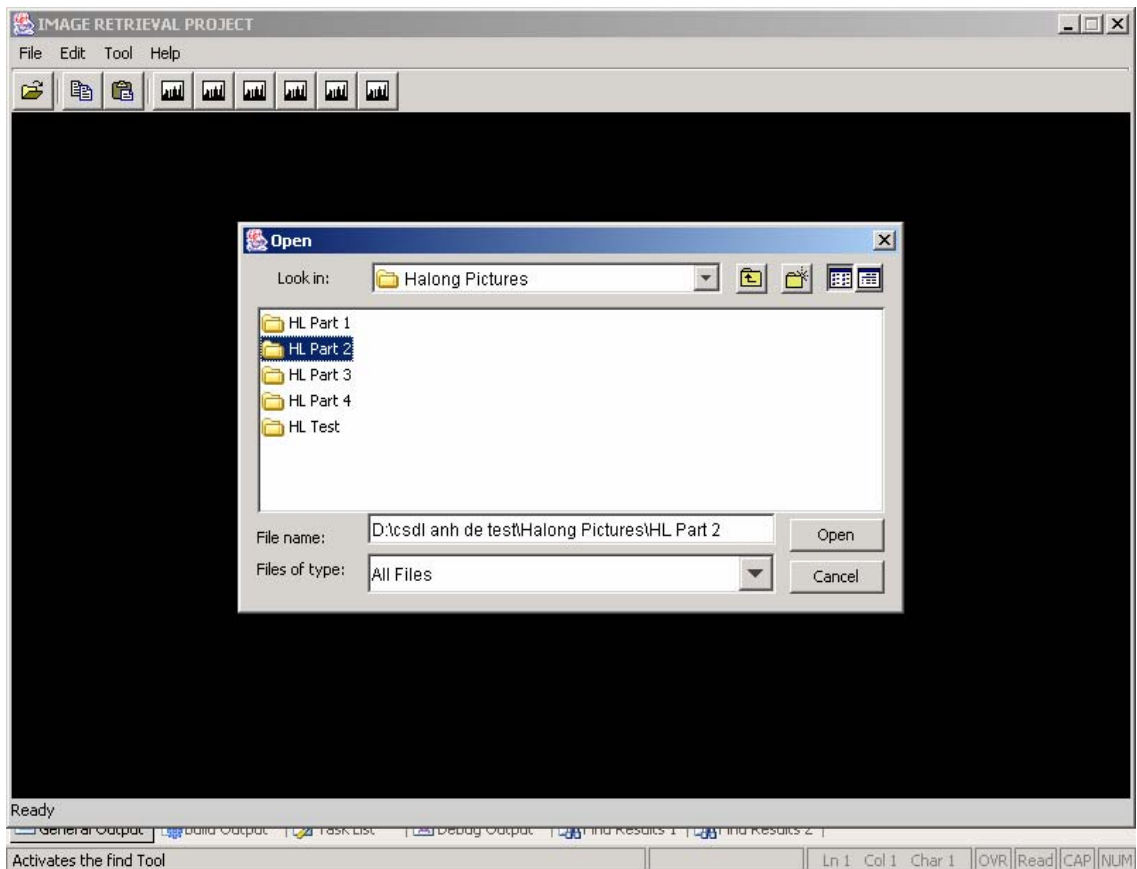
Để xem kết quả tách ra ảnh con thứ tư từ ảnh gốc, vào Tool→Region 4



Hình 4.6: Vùng 4

4.2 Rút trích các đặc trưng và lưu vào cơ sở dữ liệu

Chạy chương trình ImageGUI chọn Tool→Update Folder, rồi mở thư mục chứa các ảnh huấn luyện và chọn vùng du lịch của các ảnh trong folder này. Chương trình sẽ lần lượt rút trích các đặc trưng của tất cả các ảnh trong thư mục đã chọn và cập nhật các đặc trưng vào cơ sở dữ liệu.



Hình 4.7: Cập nhật đặc trưng vào cơ sở dữ liệu

4.3 Tìm cấu hình mạng tối ưu

Chạy chương trình myICTrainer.class, lần lượt chọn Hạ Long, Hà Nội và Nha Trang và xem kết quả tỷ lệ phân lớp chính xác và tỷ lệ bao phủ với số nút ẩn từ 1 tới 10 qua quá trình học nhanh. Qua đó chọn số nút ẩn tối ưu cho các mạng con.

Sau khi chọn số nút ẩn tối ưu, chạy file myICTrainerUserNode.class lần lượt cho Hạ Long, Hà Nội và Nha Trang để huấn luyện và lưu lại trọng số các mạng con tối ưu.

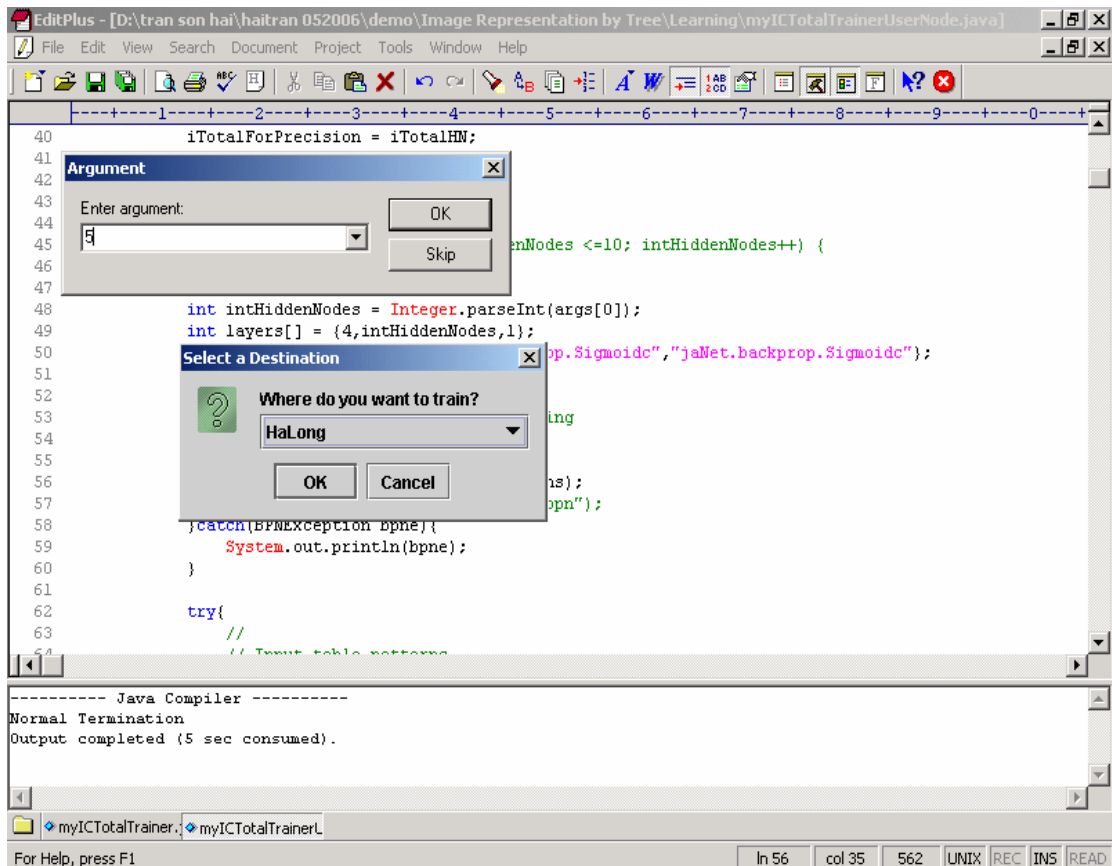
Sau khi có các mạng con tối ưu, chạy file myRegionUpdate.class cho Hạ Long, Hà Nội và Nha Trang để cập nhật output của 822 ảnh huấn luyện qua các mạng con vào cơ sở dữ liệu.

Sau khi rút trích các đặc trưng cho từng ảnh con tiến hành cho huấn luyện trên từng ảnh con theo mạng Noron lan truyền ngược. Sau khi việc huấn luyện cục bộ trên từng ảnh con hoàn tất, tiến hành tính toán và ghi nhận các output trên từng region để xác định vector đặc trưng toàn cục cho ảnh. Bây giờ mỗi ảnh huấn luyện đặc trưng bởi một vector có 4 thành phần là output của 4 ảnh con cục bộ.

Bây giờ tiến hành huấn luyện cho mạng meta-Noron toàn cục. Lần lượt cho học nhanh (chấp nhận sai số lớn) với số bước đơn điều chỉnh tham số tối đa 200000 và ghi nhận kết quả phân loại để xem thiết lập bao nhiêu nút ẩn là tối ưu. Mạng meta-Noron toàn cục sẽ quyết định mức độ quan trọng của các mạng cục bộ nên sẽ ảnh hưởng lớn đến kết quả truy vấn.

Chạy chương trình myICTotalTrainer.class, lần lượt chọn Hạ Long, Hà Nội và Nha Trang và xem kết quả tỷ lệ phân lớp chính xác và tỷ lệ bao phủ với số nút ẩn từ 1 tới 10 qua quá trình học nhanh. Qua đó chọn số nút ẩn tối ưu cho các mạng meta-Noron toàn cục .

Sau khi chọn số nút ẩn tối ưu, chạy file myICTotal TrainerUserCode.class lần lượt cho Hạ Long, Hà Nội và Nha Trang để huấn luyện và lưu lại trọng số các mạng meta-Noron toàn cục tối ưu.



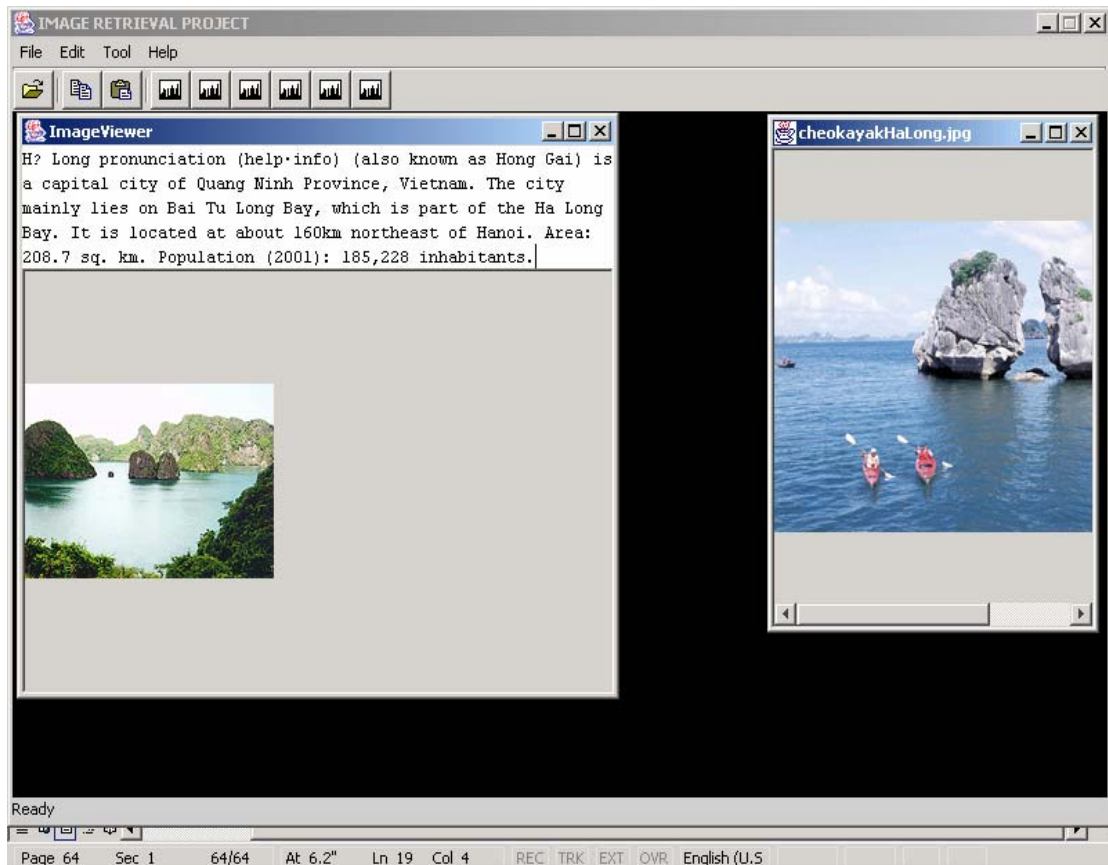
Hình 4.8: Huấn luyện với số nút ẩn tối ưu

Sau khi có bộ trọng số mạng toàn cục tối ưu, chạy file `myTotalUpdateOutput.class` cho Hạ Long, Hà Nội và Nha Trang để cập nhật output của 822 ảnh huấn luyện qua các mạng meta-Noron toàn cục vào cơ sở dữ liệu.

Trong phạm vi nghiên cứu này tiến hành trên tập 822 ảnh huấn luyện gồm 201 ảnh thuộc về Hạ Long, 367 ảnh thuộc về Hà Nội và 254 ảnh thuộc về Nha Trang. Căn cứ vào kết quả thực nghiệm trên thì nên chọn số nút ẩn cho mạng meta-Noron Hạ Long là 10, cho Hà Nội là 8 và cho Nha Trang là 8 nút ẩn là tối ưu về tổng thể dựa trên ba yếu tố: độ chính xác, độ bao phủ và thời gian huấn luyện.

4.4 Thực hiện truy vấn tìm vùng du lịch

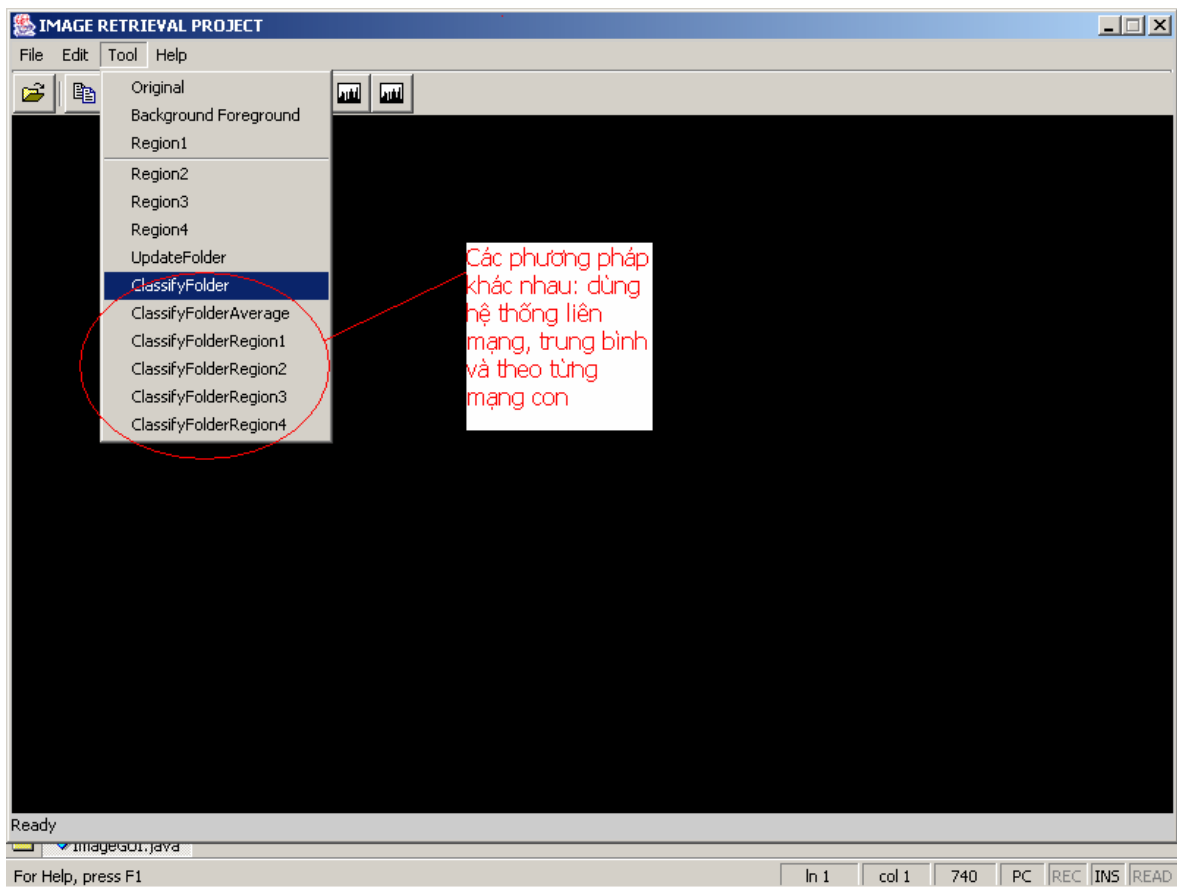
Chạy chương trình ImageGUI.class rồi chọn File→Open mở một file ảnh dùng để truy vấn nào đó. Sau đó Vào Edit→Searching, hệ thống sẽ tìm vùng du lịch và hiển thị ảnh của vùng du lịch đó



Hình 4.9: Kết quả truy tìm vùng du lịch

Tiến hành truy vấn kiểm thử trên 82 ảnh được kết quả sau tìm đúng vùng du lịch là 59 ảnh đạt tỷ lệ: 71.95%. Để nâng cao hơn nữa hiệu quả truy vấn vùng du lịch cần sử dụng hệ thống rút trích các đặc trưng trong ảnh tốt hơn của một số tác giả khác đã và đang nghiên cứu chuyên về rút trích đặc trưng. Trong phạm vi luận văn này, tập trung chủ yếu vào việc xây dựng và phát triển hệ thống liên mạng meta-Noron.

Chạy file ImageGUI.class, lần lượt chọn các phương pháp khác nhau:



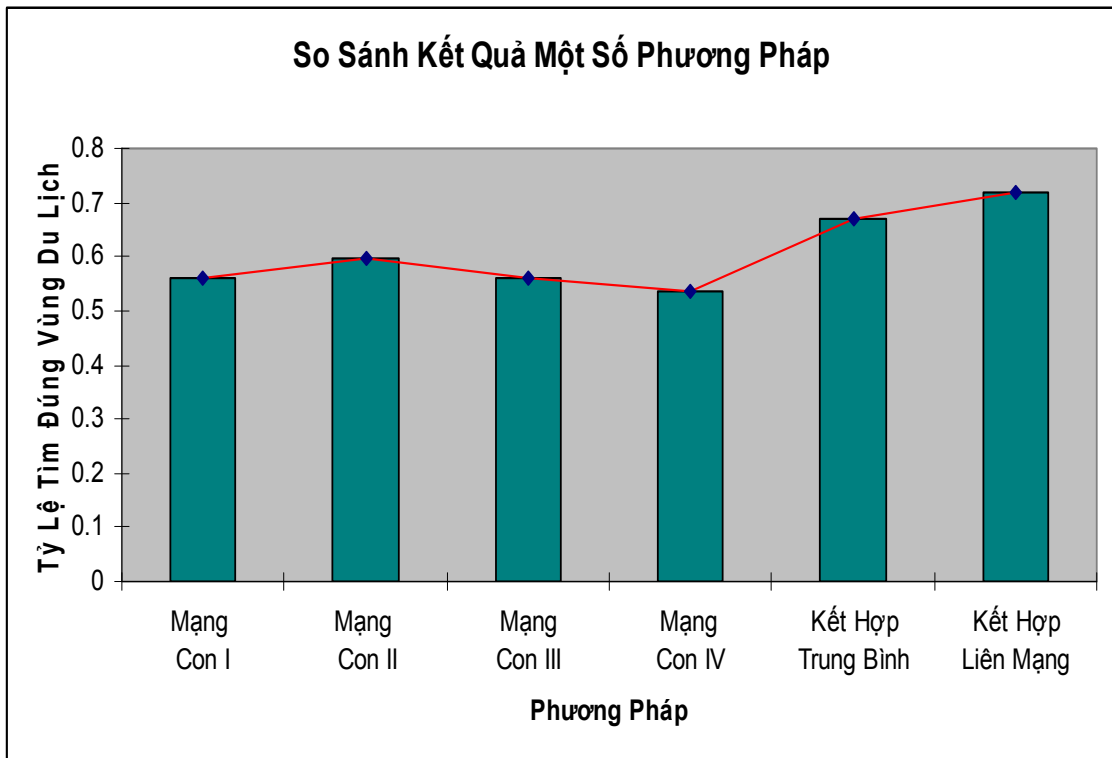
Hình 4.10: Chức năng thống kê xác định vùng du lịch

Tiến hành truy vấn tìm vùng du lịch trên 82 ảnh test theo một số phương pháp: dựa vào một mạng Noron con, kết hợp trung bình (đánh đồng mức độ tin cậy của 4 vùng) và dùng hệ thống liên mạng được bảng kết quả sau:

Bảng 4.1: Kết quả thực nghiệm một số phương pháp

	Mạng Con I	Mạng Con II	Mạng Con III	Mạng Con IV	Kết Hợp Trung Bình	Kết Hợp Liên Mạng
Hạ Long	13	13	14	12	16	15
Hà Nội	17	21	22	20	24	22
Nha Trang	16	15	10	12	15	22
Tổng Cộng	46	49	46	44	55	59

Từ đó vẽ ra biểu đồ so sánh kết quả sau:



Hình 4.11: Biểu đồ so sánh một số phương pháp

Rõ ràng chỉ dựa trên đánh giá của từng mạng con mà truy vấn thì tỷ lệ truy tìm chính xác vùng du lịch không cao, khi kết hợp kết quả các mạng con lại dù là phương pháp trung bình hay dùng mạng Nơron kết hợp thì hiệu quả đều được nâng lên đáng kể. Tuy nhiên, phương pháp dùng mạng Nơron kết hợp các kết quả sau quá trình huấn luyện sẽ cho đánh giá mức độ tin cậy của các mạng con chính xác hơn phương pháp trung bình.

Chương 5: KẾT LUẬN

Truy vấn ảnh dựa trên nội dung (CBIR) đến nay vẫn là lĩnh vực nghiên cứu rất đáng quan tâm và có nhiều phương pháp và cách tiếp cận khác nhau. Hầu hết các ứng dụng hữu ích của CBIR hiện nay là các ứng dụng áp dụng trực tiếp kỹ thuật CBIR ở mức 1. Như các ứng dụng nhằm ngăn ngừa tội phạm, hệ thống tìm kiếm vân tay, các hệ thống nhận dạng khuôn mặt hay các ứng dụng trong quân đội: tìm kiếm có máy bay trong màn hình rada hay không, nếu có là máy bay gì loại nào; các hệ thống hỗ trợ quản lý đăng ký logo (xác định xem logo muốn đăng ký là hoàn toàn mới hay đã có logo tương tự đăng ký trước đó)...

Tính khoa học của luận văn thể hiện qua phương pháp truy vấn ảnh dùng hệ thống liên mạng meta-Noron. Mạng Noron con với đầu vào là các vector đặc trưng dạng lượng tử gồm đặc trưng gồm thông tin về màu sắc, hình dạng và vị trí đã đề xuất, sẽ kết hợp các đặc trưng này cho ra kết quả đánh giá của mạng con. Sau đó, để nâng cao độ chính xác của hệ thống mạng meta-Noron toàn cục sẽ kết hợp kết quả của các mạng con cho ra kết quả cuối cùng của toàn hệ thống liên mạng. Xây dựng hệ thống truy vấn ảnh, đồng thời chỉ ra tính khả thi của hệ thống so với các hệ thống khác qua các kết quả thực nghiệm. Giá trị thực tiễn của luận văn được chứng minh qua ứng dụng thực tế tra cứu vùng du lịch.

Luận văn này đã đề xuất một phương pháp kết hợp các đặc trưng màu sắc, hình dạng và vị trí bằng hệ thống liên mạng meta-Noron. Hệ thống liên mạng meta-Noron không những áp dụng vào các đặc trưng màu sắc, hình dạng và vị trí của ảnh mà còn có thể áp dụng cho các đặc trưng lượng tử khác. Các đặc trưng lượng tử bất kỳ rút trích từ ảnh sẽ qua các mạng Noron con cho ra các output cục bộ. Các output của các mạng con sẽ là ngõ vào của mạng meta-Noron để cho ra các kết quả output cuối cùng của toàn hệ thống liên mạng.

Tuy nhiên, luận văn vẫn còn những hạn chế nhất định: cần hoàn thiện ứng dụng tra cứu vùng du lịch và đề ra phương pháp để tự động xác định cấu hình hệ thống mạng Nơron tối ưu cho lớp bài toán tương ứng.

Kết hợp hệ thống liên mạng meta-Nơron với các phương pháp rút trích đối tượng trong ảnh đang nghiên cứu của các tác giả khác để nâng cao kết quả truy vấn của hệ thống đã đề xuất là một hướng phát triển của luận văn này.

TÀI LIỆU THAM KHẢO

Tiếng Việt

- [1] Lâm Thị Ngọc Châu (2003), *"Xây dựng chỉ mục cho Cơ Sở Dữ Liệu Ảnh"*, Luận văn thạc sĩ của khoa Công Nghệ Thông Tin, Trường Đại học Khoa Học Tự Nhiên, Tp. Hồ Chí Minh
- [2] Nguyễn Văn Diêu (2003), *"Biến đổi wavelet và ứng dụng trong tìm kiếm trên Cơ Sở Dữ Liệu Ảnh"*, Luận văn thạc sĩ của khoa Công Nghệ Thông Tin, Trường Đại học Khoa Học Tự Nhiên, Tp. Hồ Chí Minh
- [3] Hoàng Kiếm, Lê Hoài Bắc, Lê Hoàng Thái (2000), *"Mạng Noron và thuật giải di truyền: ứng dụng cho nhận dạng chữ viết tay"*, Hội nghị khoa học lần thứ hai: Báo cáo khoa học Công Nghệ Thông Tin, Trường Đại học Khoa Học Tự Nhiên, Tp. Hồ Chí Minh.
- [4] Nguyễn Phước Lộc, *"Truy vấn Cơ Sở Dữ Liệu ảnh qua mô hình Mediator Markov Model"*, Luận văn thạc sĩ Công Nghệ Thông Tin, Trường Đại học Khoa Học Tự Nhiên, Tp. Hồ Chí Minh.
- [5] Lê Hoàng Thái (2004), *"Xây Dựng, Phát Triển, Ứng Dụng Một Số Mô Hình Kết Hợp Giữa Mạng Noron (NN), Logic Mờ (FL) và Thuật Giải Di Truyền (GA)"*, Luận án Tiến Sĩ Toán học, Trường Đại học Khoa Học Tự Nhiên, Tp. Hồ Chí Minh.

Tiếng Anh

- [6] Lê Hoài Bắc, Lê Hoàng Thái (2004), *"The GA_NN_FL associated model for authenticating fingerprinter"*, the KES'2004 International Program Committee, Wellington Institute of Technology, NEW ZEALAND.

- [7] Brown, R. and B. Pham (2005), "*Image Mining and Retrieval Using Hierarchical Support Vector Machines*", Proceedings of the 11th International Multimedia Modelling Conference (MMM'05)-Volume 00, p. 446-451.
- [8] Castelli, V., et al. (1998), "*Progressive search and retrieval in large image archives*", IBM J RES DEV, 42(2): p. 253-268.
- [9] Chen, Y. and J.Z. Wang (2002), "*A region-based fuzzy feature matching approach to content-based image retrieval*", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(9): p. 1252-1267.
- [10] Eakins, J., M. Graham, and N. University of Northumbria at (1999), "*Content-based Image Retrieval*", University of Northumbria at Newcastle.
- [11] Ghoshal, A., P. Ircing, and S. Khudanpur (2005), "*Hidden Markov models for automatic annotation and content-based retrieval of images and video*", Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, p. 544-551.
- [12] Hoiem, D., et al. (2004), "*Object-based image retrieval using the statistical structure of images*", Computer Vision and Pattern Recognition, CVPR 2004. Proceedings of the IEEE Computer Society Conference on. 2.
- [13] Lee, H.K. and S. Yoo (2001), "*A Neural Network-based Flexible Image Retrieval*"
- [14] Prasad, B.G., K.K. Biswas, and S.K. Gupta (2004), "*Region-based image retrieval using integrated color, shape, and location index*", Computer Vision and Image Understanding, 94(1-3): p. 193-233.

- [15] Schettini, R., G. Ciocca, and S. Zuffi (2001), "*A survey on methods for colour image indexing and retrieval in image databases*", Color Imaging Science: Exploiting Digital Media, p. 183-211.
- [16] Tong, S. and E. Chang (2001), "*Support vector machine active learning for image retrieval*", Proceedings of the ninth ACM international conference on Multimedia, p. 107-118.
- [17] Torres, J.M. and A.P. Parkes (2000), "*User modelling and adaptivity in visual information retrieval systems*", Workshop on Computational Semiotics for New Media, University of Surrey, Surrey, UK, June of, 2000.
- [18] Van Den Broek, E.L., P.M.F. Kisters, and L.G. Vuurpijl (2005), "*Content-based image retrieval benchmarking: Utilizing color categories and color distributions*", The Journal of imaging science and technology, 49(3): p. 293-301.
- [19] Van Rikxoort, E.M., E.L. Van den Broek, and T.E. Schouten (2004), "*the development of a human-centered object-based image retrieval engine*", Proceedings of the Eleventh Annual Conference of the Advanced School for Computing and Imaging: p. 401–408.
- [20] Veltkamp, R.C. and M. Tanase (2000), "*Content-based image retrieval systems: A survey*", Utrecht, Netherlands: Department of Computing Science, Utrecht University.
- [21] Vertan, C. and N. Boujemaa (2000), "*Embedding fuzzy logic in content based image retrieval*", Fuzzy Information Processing Society, 2000. NAFIPS. 19th International Conference of the North American, p. 85-89.

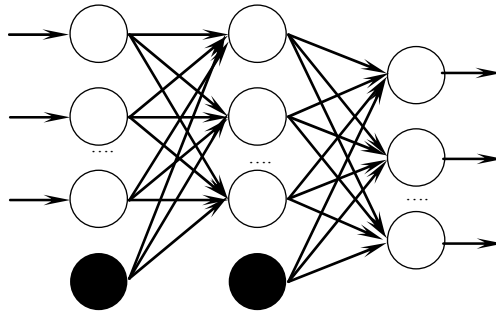
- [22] Zhang, L., F. Lin, and B. Zhang (2001), “*Support vector machine learning for image retrieval*”, Image Processing, Proceedings International Conference on, 2.

Websites

- [23] <http://debut.cis.nctu.edu.tw/pages/slides/jeffrey/present2.pdf> (Oct 2005)
- [24] <http://homepages.inf.ed.ac.uk/rbf/CVOnline/dbase.htm> (Oct 2005)
- [25] Blobworld, <http://elib.cs.berkeley.edu/photos/blobworld/> (Nov 2005)
- [26] AltaVista Photofinder, <http://image.altavista.com/cgi-bin/avn CGI>.
(Jan 2006)
- [27] ImageMiner, <http://www.tzi.de/bv/ImageMinerhtml/> (Jan 2006)
- [28] NETRA, <http://maya.ece.ucsb.edu/Netra/> (Jan 2006)
- [29] Query By Image Content, <http://wwwqbic.almaden.ibm.com/> (Jan 2006)
- [30] VisualSeek, <http://www.ctr.columbia.edu/VisualSEEk/> (Feb 2006)
- [31] WebSeek, <http://www.ctr.columbia.edu/WebSEEk/> (Feb 2006)
- [32] Wavelet Image Search Engine,
<http://www.stanford.edu/~wangz/project/imsearch/> (Feb 2006)
- [33] Remco C. Veltkamp, Mirela Tanase, Department of Computing Science, Utrecht University, “*Content-Based Image Retrieval Systems: A Survey*”,
<http://give-lab.cs.uu.nl/cbirsurvey/cbir-survey/index.html> (Mar 2006)

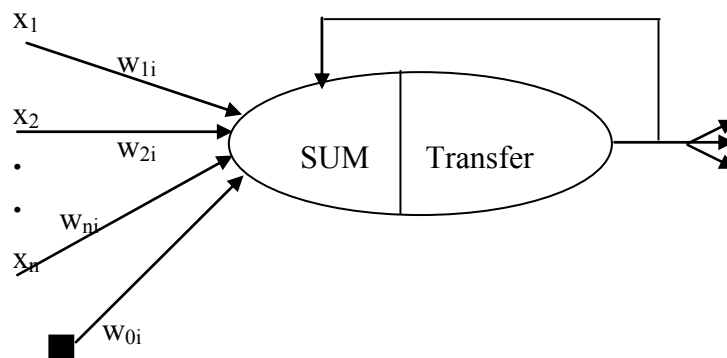
PHỤ LỤC A: Mạng Nơron Lan Truyền Ngược

A.1 Cấu Trúc Mạng



Hình A.1: Mạng Nơron

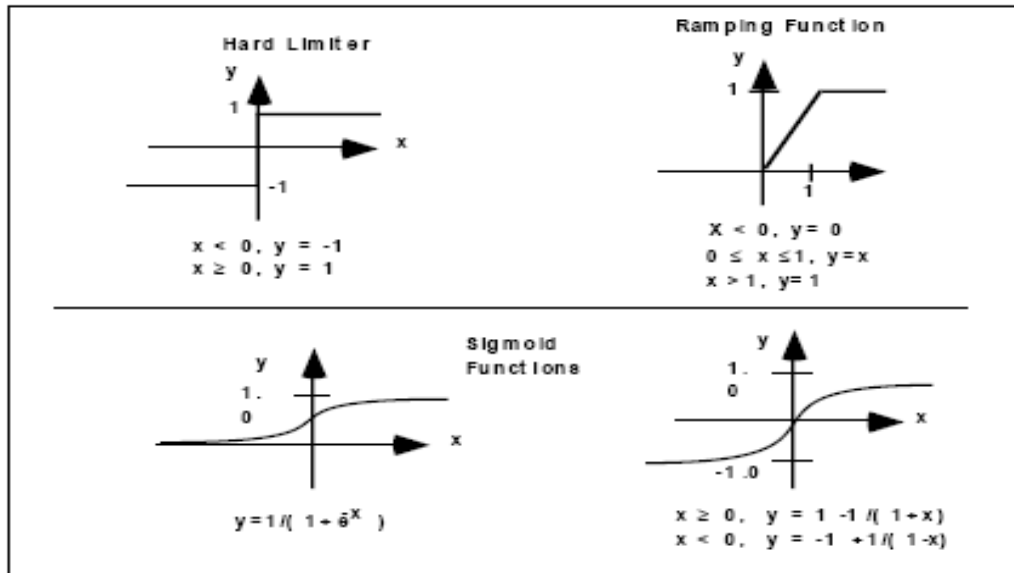
Hình A.1 trên minh họa một mạng Nơron gồm 3 lớp: 1 lớp input, 1 lớp ẩn và 1 lớp output. Mỗi lớp gồm nhiều đơn vị liên kết với các đơn vị của lớp kế tiếp bởi các trọng số. Lớp ẩn chứa các Nơron, mỗi Nơron có cơ chế xử lý thông tin: nối kết và biến đổi toán học bên trong Nơron. Hoạt động của một Nơron trong trường hợp mạng có 1 lớp ẩn có thể minh họa như hình sau:



Hình A.2: Hoạt động của một Nơron

Từ các nút nhập x , dựa vào hàm tổng (Summation) tính ra giá trị các nút ẩn trung gian H . Sau đó, cho giá trị H qua hàm truyền (Transfer) để tính ra các giá trị xuất $Y = f(H)$ với f là hàm truyền.

Có một số hàm truyền thông dụng như: Limiter, Rampling và Sigmod



Hình A.3: Một số hàm truyền

A.2 Thuật giải lan truyền ngược

Cho không gian các mẫu học (x, t) , x là giá trị cần huấn luyện và t là giá trị output kỳ vọng của quá trình huấn luyện, hệ số học α . Thuật giải lan truyền ngược được tóm tắt như sau:

1. Tạo mạng truyền thẳng gồm n_{in} nút input, h lớp ẩn mỗi lớp ẩn có n_{hidden} nút và n_{out} nút output.
2. Khởi tạo bộ trọng số cho mạng.
3. Trong khi <điều kiện dừng chưa thỏa> làm

Với mỗi cặp (x, t) trong không gian mẫu huấn luyện thực hiện

- I. Xét lớp ẩn: truyền x qua mạng, tại mỗi lớp xác định đầu ra của mỗi Noron. Quá trình này thực hiện cho đến khi gặp lớp xuất.

- II. Xét lớp xuất: đối với đầu ra O_k của Nơron k trong lớp xuất K , xác định sai số δ_k của nó: $\delta_k = O_k(1 - O_k)(t_k - O_k)$ (A.1). Chuyển sang lớp ẩn $L = K - 1$.
- III. Xét lớp ẩn: với mỗi Nơron trên lớp ẩn thứ L , xác định sai số δ_t của nó

$$\delta_t = O_t(1 - O_t) \sum_{i \in L+1} w_{it} \delta_i \quad (\text{A.2})$$

- IV. Cập nhật trọng số trong mạng, $w_{hj} = w_{hj} + \alpha * \delta_{oj} * o_j$ (A.3)
- V. Nếu $(L > 1)$ thì chuyển sang lớp ẩn kế tiếp và quay về bước III
- ngược lại thì chọn cặp mẫu học khác và quay lại bước I

Tóm lại, giá trị sai số tính đầu tiên cho các Nơron lớp xuất, kết quả này được dùng để tính sai số của các Nơron lớp ẩn cao nhất, các lớp ẩn kế tiếp thì tính dựa theo kết quả của lớp ẩn cao hơn. Cứ thế lan truyền ngược cho đến lớp nhập.

PHỤ LỤC B: Cơ Sở Dữ Liệu Ảnh

Tập ảnh huấn luyện gồm 822 ảnh, trong đó có 201 ảnh Hạ Long, 367 ảnh Hà Nội và 254 ảnh Nha Trang.

Tập ảnh kiểm tra gồm 82 ảnh, trong đó có 22 ảnh Hạ Long, 35 ảnh Hà Nội và 25 ảnh Nha Trang

B.1 Một số ảnh huấn luyện



Hình B.1: Ảnh huấn luyện 1



Hình B.2: Ảnh huấn luyện 2



Hình B.3: Ảnh huấn luyện 3



Hình B.4: Ảnh huấn luyện 4

B.2 Một số ảnh kiểm tra



Hình B.5: Ảnh kiểm tra 1



Hình B.6: Ảnh kiểm tra 2



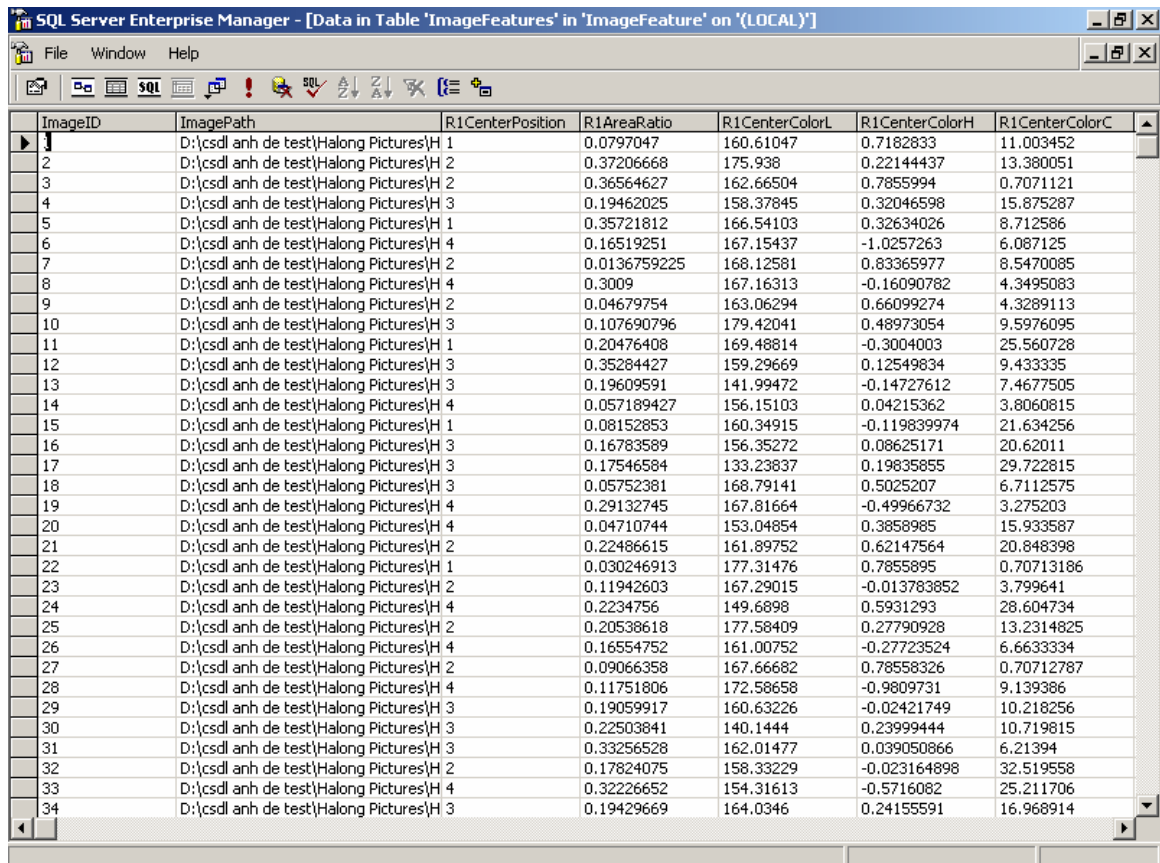
Hình B.7: Ảnh kiểm tra 3



Hình B.8: Ảnh kiểm tra 4

PHỤ LỤC C: Cơ Sở Dữ Liệu Đặc Trưng

Các đặc trưng của ảnh được lưu trong bảng ImageFeatures trên database SQL Server 2000



ImageID	ImagePath	R1CenterPosition	R1AreaRatio	R1CenterColorL	R1CenterColorH	R1CenterColorC
1	D:\csdl anh de test\Halong Pictures\H 1	0.0797047	160.61047	0.7182833	11.003452	
2	D:\csdl anh de test\Halong Pictures\H 2	0.37206668	175.938	0.22144437	13.380051	
3	D:\csdl anh de test\Halong Pictures\H 2	0.36564627	162.66504	0.7855994	0.7071121	
4	D:\csdl anh de test\Halong Pictures\H 3	0.19462025	158.37845	0.32046598	15.875287	
5	D:\csdl anh de test\Halong Pictures\H 1	0.35721812	166.54103	0.32634026	8.712586	
6	D:\csdl anh de test\Halong Pictures\H 4	0.16519251	167.15437	-1.0257263	6.087125	
7	D:\csdl anh de test\Halong Pictures\H 2	0.0136759225	168.12581	0.83365977	8.5470085	
8	D:\csdl anh de test\Halong Pictures\H 4	0.3009	167.16313	-0.16090782	4.3495083	
9	D:\csdl anh de test\Halong Pictures\H 2	0.04679754	163.06294	0.66099274	4.3289113	
10	D:\csdl anh de test\Halong Pictures\H 3	0.107690796	179.42041	0.48973054	9.5976095	
11	D:\csdl anh de test\Halong Pictures\H 1	0.20476408	169.48814	-0.3004003	25.560728	
12	D:\csdl anh de test\Halong Pictures\H 3	0.35284427	159.29669	0.12549834	9.433335	
13	D:\csdl anh de test\Halong Pictures\H 3	0.19609591	141.99472	-0.14727612	7.4677505	
14	D:\csdl anh de test\Halong Pictures\H 4	0.057189427	156.15103	0.04215362	3.8060815	
15	D:\csdl anh de test\Halong Pictures\H 1	0.08152853	160.34915	-0.119839974	21.634256	
16	D:\csdl anh de test\Halong Pictures\H 3	0.16783589	156.35272	0.08625171	20.62011	
17	D:\csdl anh de test\Halong Pictures\H 3	0.17546584	133.23837	0.19835855	29.722815	
18	D:\csdl anh de test\Halong Pictures\H 3	0.05752381	168.79141	0.5025207	6.7112575	
19	D:\csdl anh de test\Halong Pictures\H 4	0.29132745	167.81664	-0.49966732	3.275203	
20	D:\csdl anh de test\Halong Pictures\H 4	0.04710744	153.04854	0.3858985	15.933587	
21	D:\csdl anh de test\Halong Pictures\H 2	0.22486615	161.89752	0.62147564	20.848398	
22	D:\csdl anh de test\Halong Pictures\H 1	0.030246913	177.31476	0.7855895	0.70713186	
23	D:\csdl anh de test\Halong Pictures\H 2	0.11942603	167.29015	-0.013783852	3.799641	
24	D:\csdl anh de test\Halong Pictures\H 4	0.2234756	149.6898	0.5931293	28.604734	
25	D:\csdl anh de test\Halong Pictures\H 2	0.20538618	177.58409	0.27790928	13.2314825	
26	D:\csdl anh de test\Halong Pictures\H 4	0.16554752	161.00752	-0.27723524	6.6633334	
27	D:\csdl anh de test\Halong Pictures\H 2	0.09066358	167.66682	0.78558326	0.70712787	
28	D:\csdl anh de test\Halong Pictures\H 4	0.11751806	172.58658	-0.9809731	9.139386	
29	D:\csdl anh de test\Halong Pictures\H 3	0.19059917	160.63226	-0.02421749	10.218256	
30	D:\csdl anh de test\Halong Pictures\H 3	0.22503841	140.1444	0.23999444	10.719815	
31	D:\csdl anh de test\Halong Pictures\H 3	0.33256528	162.01477	0.039050866	6.21394	
32	D:\csdl anh de test\Halong Pictures\H 2	0.17824075	158.33229	-0.023164898	32.519558	
33	D:\csdl anh de test\Halong Pictures\H 4	0.32226652	154.31613	-0.5716082	25.211706	
34	D:\csdl anh de test\Halong Pictures\H 3	0.19429669	164.0346	0.24155591	16.968914	

Hình C.1: Cơ sở dữ liệu đặc trưng

Cấu trúc của bảng dữ liệu ImageFeatures như sau:

Bảng C.1: Mô tả cơ sở dữ liệu đặc trưng

Nhóm	Tên Cột	Ý nghĩa
Nhóm mở rộng	ImageID	Mã Khóa chính
	ImagePath	Đường dẫn file hình
Nhóm các đặc trưng của vùng 1	R1CenterPosition	Đặc trưng thứ nhất của vùng 1 của ảnh
	R1AreaRatio	Đặc trưng thứ hai của vùng 1 của ảnh
	R1CenterColorL	Đặc trưng thứ ba của vùng 1 của ảnh
	R1CenterColorH	Đặc trưng thứ tư của vùng 1 của ảnh
	R1CenterColorC	Đặc trưng thứ năm của vùng 1 của ảnh
Nhóm các đặc trưng của vùng 2	R2CenterPosition	Đặc trưng thứ nhất của vùng 2 của ảnh
	R2AreaRatio	Đặc trưng thứ hai của vùng 2 của ảnh
	R2CenterColorL	Đặc trưng thứ ba của vùng 2 của ảnh
	R2CenterColorH	Đặc trưng thứ tư của vùng 2 của ảnh
	R2CenterColorC	Đặc trưng thứ năm của vùng 2 của ảnh
Nhóm các đặc trưng của vùng 3	R3CenterPosition	Đặc trưng thứ nhất của vùng 3 của ảnh

	R3AreaRatio	Đặc trưng thứ hai của vùng 3 của ảnh
	R3CenterColorL	Đặc trưng thứ ba của vùng 3 của ảnh
	R3CenterColorH	Đặc trưng thứ tư của vùng 3 của ảnh
	R3CenterColorC	Đặc trưng thứ năm của vùng 3 của ảnh
Nhóm các đặc trưng của vùng 4	R4CenterPosition	Đặc trưng thứ nhất của vùng 4 của ảnh
	R4AreaRatio	Đặc trưng thứ hai của vùng 4 của ảnh
	R4CenterColorL	Đặc trưng thứ ba của vùng 4 của ảnh
	R4CenterColorH	Đặc trưng thứ tư của vùng 4 của ảnh
	R4CenterColorC	Đặc trưng thứ năm của vùng 4 của ảnh
Nhóm mở rộng	Place	Vùng du lịch
Nhóm các output mạng con 1	R1HaLongLocalOutput	Kết quả nút output thứ nhất của mạng con 1
	R1HaNoiLocalOutput	Kết quả nút output thứ hai của mạng con 1
	R1NhaTrangLocalOutput	Kết quả nút output thứ ba của mạng con 1
Nhóm các output mạng con 2	R2HaLongLocalOutput	Kết quả nút output thứ nhất của mạng con 2

	R2HaNoiLocalOutput	Kết quả nút output thứ hai của mạng con 2
	R2NhaTrangLocalOutput	Kết quả nút output thứ ba của mạng con 2
Nhóm các output mạng con 3	R3HaLongLocalOutput	Kết quả nút output thứ nhất của mạng con 3
	R3HaNoiLocalOutput	Kết quả nút output thứ hai của mạng con 3
	R3NhaTrangLocalOutput	Kết quả nút output thứ ba của mạng con 3
Nhóm các output mạng con 4	R4HaLongLocalOutput	Kết quả nút output thứ nhất của mạng con 4
	R4HaNoiLocalOutput	Kết quả nút output thứ hai của mạng con 4
	R4NhaTrangLocalOutput	Kết quả nút output thứ ba của mạng con 4
Nhóm output toàn cục	HaLongTotalOut	Kết quả nút output thứ nhất toàn cục
	HaNoiTotalOut	Kết quả nút output thứ hai toàn cục
	NhaTrangTotalOut	Kết quả nút output thứ ba toàn cục
Nhóm mở rộng	ContextText	Thông tin chi tiết về vùng du lịch

Lưu các thông tin: tên file ảnh, vùng du lịch và thông tin chi tiết của 822 ảnh huấn luyện vào các cột tương ứng ImagePath, Place và ContextText.

Rút trích các đặc trưng của vùng 1 của ảnh vào các nhóm cột đặc trưng của vùng 1. Rút trích các đặc trưng của vùng 2 của ảnh vào các nhóm cột đặc trưng của vùng 2. Rút trích các đặc trưng của vùng 3 của ảnh vào các nhóm cột đặc trưng của vùng 3. Rút trích các đặc trưng của vùng 4 của ảnh vào các nhóm cột đặc trưng của vùng 4.

Sử dụng giá trị của tất cả 822 dòng tại nhóm cột đặc trưng vùng 1 và Place để huấn luyện cho mạng con thứ nhất RN_1 . Huấn luyện xong ghi nhận output của mạng con này vào nhóm cột các output của mạng con 1. Tiến hành tương tự cho các mạng con RN_2, RN_3, RN_4 .

Sử dụng giá trị của tất cả 822 dòng ứng với các cột R1HaLongLocalOutput, R2HaLongLocalOutput, R3HaLongLocalOutput, R4HaLongLocalOutput và Place để huấn luyện mạng thành phần TP_1 . Sau khi huấn luyện xong mạng thành phần này lưu giá trị output của mạng thành phần vào cột HaLongTotalOut. Tương tự cho các mạng thành phần TP_2 , và TP_3 .