

Bài báo nghiên cứu

TÌM KIẾM ẢNH SỬ DỤNG MẠNG NƠN TÍCH CHẬP VÀ ĐỒ THỊ PHÂN CỤM

Phạm Hoàng Phương¹, Đỗ Xuân Hiệp², Nguyễn Thị Định³, Văn Thế Thành^{4*}

¹Trường THCS Bình Tân, Lagi, Bình Thuận, Việt Nam

²Bảo hiểm xã hội tỉnh Bà Rịa – Vũng Tàu, Việt Nam

³Trường Đại học Công nghiệp Thực phẩm Thành phố Hồ Chí Minh, Việt Nam

⁴Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ: Văn Thế Thành – Email: thanhvt@hcmue.edu.vn

Ngày nhận bài: 11-10-2022; ngày nhận bài sửa: 05-11-2022; ngày duyệt đăng: 26-6-2023

TÓM TẮT

Trong bài báo này, một mô hình tìm kiếm ảnh dựa trên mạng nơon tích chập kết hợp cấu trúc đồ thị cụm được thực hiện nhằm nâng cao hiệu suất và giảm thời gian truy vấn ảnh. Để thực hiện bài toán này: (1) mạng Noron tích chập được sử dụng để xác định và phân loại các đối tượng trên ảnh; (2) cấu trúc đồ thị cụm được xây dựng để thực hiện xây dựng ontology; (3) tập ảnh tương tự được trích xuất dựa trên ontology sau thực hiện khi tìm kiếm bằng câu truy vấn SPARQL. Với mỗi ảnh đầu vào, sau khi phân loại từng đối tượng bằng mạng Noron tích chập; trích xuất vector đặc trưng; phân lớp ảnh và thực hiện truy vấn ontology để trích xuất tập ảnh tương tự. Trên cơ sở lý thuyết đề xuất, một mô hình truy vấn ảnh được đề xuất và thực nghiệm trên bộ ảnh COCO, Flickr với độ chính xác tương ứng lần lượt là 0.7950, 0.8116. Theo kết quả thực nghiệm, phương pháp đề xuất của chúng tôi được đánh giá là đúng đắn và so sánh với các công trình khác trên cùng bộ ảnh nhằm đánh giá tính hiệu quả của mô hình đề xuất; đồng thời áp dụng được cho các bộ dữ liệu khác nhau.

Từ khóa: mạng nơon tích chập; tìm kiếm ảnh; ảnh tương tự; SPARQL

1. Giới thiệu

Tìm kiếm ảnh tương tự là bài toán được ứng dụng trong nhiều lĩnh vực như nhận dạng mẫu, nhận dạng ký tự, phân loại bệnh nhân và chẩn đoán y khoa qua hình ảnh... nhằm tận dụng nguồn dữ liệu đa phương tiện đang phát triển nhanh chóng trong bối cảnh hiện nay (Kumar, 2016; Ptucha, 2019). Dữ liệu ảnh số tăng lên theo cấp số nhân trong mỗi giây thông qua các thiết bị đa phương tiện; điều này cho thấy sự cần thiết phải có một phương pháp tra cứu ảnh để nâng cao hiệu suất tìm kiếm ảnh trong khoảng thời gian cho phép để đáp ứng nhu cầu người dùng. Bài toán tìm kiếm ảnh tương tự là một trong những bài toán quan trọng

Cite this article as: Phạm Hoàng Phương, Đỗ Xuân Hiệp, Nguyễn Thị Định, & Văn Thế Thành (2023). Image retrieval using convolutional neural networks and cluster graph. *Ho Chi Minh City University of Education Journal of Science*, 20(7), 1141-1154.

của nhiều hệ tra cứu dữ liệu đa phương tiện được nhiều nhóm nghiên cứu quan tâm (Ptucha, 2019). Để thực hiện bài toán tìm kiếm ảnh tương tự cần xây dựng một mô hình và đánh giá tính đúng đắn của mô hình đề xuất. Hiện nay, có nhiều phương pháp để nâng cao hiệu quả tìm kiếm ảnh bằng cách sử dụng nhiều kỹ thuật khác nhau hoặc kết hợp nhiều kỹ thuật cho mỗi bài toán. Vì vậy, trong bài báo này, một phương pháp kết hợp mạng nơ-ron tích chập kết hợp với cấu trúc đồ thị cụm để thực hiện bài toán truy vấn ảnh đã mang lại kết quả khả quan.

Đóng góp của bài báo gồm: (1) Trích xuất đặc trưng thị giác của hình ảnh dựa trên mạng tích chập; (2) xây dựng cấu trúc đồ thị cụm dựa trên thuật toán K-Mean; (3) phân bố các hình ảnh trên ontology và thực hiện truy vấn ảnh tương tự với ảnh đầu vào; (4) đề xuất mô hình tìm kiếm ảnh dựa trên mạng nơ-ron tích chập và ontology; xây dựng thực nghiệm và chứng minh tính đúng đắn, hiệu quả của phương pháp đề xuất dựa trên các bộ dữ liệu ảnh đa đối tượng MS-COCO (MS-COCO, 2017), Flickr (Flickr, 2017).

Các công trình nghiên cứu liên quan

Phương pháp gom cụm đã được ứng dụng trong nhiều hệ thống tìm kiếm ảnh như áp dụng thuật toán K-means để gom cụm đặc trưng màu sắc hình ảnh (Lin et al., 2014); tìm kiếm ảnh dựa trên thuật toán K-means và khoảng cách Mahalanobis giữa các véc-tơ màu sắc của các hình ảnh (Cevikalp et al., 2018); áp dụng thuật toán K-means và dải màu MPEG7 cho việc gom cụm và tìm kiếm ảnh (Saboorian et al., 2010); thực hiện gom cụm kết hợp đặc trưng màu sắc, hình dạng và cấu trúc hình ảnh để tìm kiếm ảnh tương tự... Ngoài ra, nhiều công trình nghiên cứu về tìm kiếm ảnh tương tự dựa trên ontology cũng đã được giới thiệu trong những năm gần đây như: tìm kiếm ảnh dựa trên từ thị giác (Jabeen et al., 2018), tra cứu ảnh dựa trên việc phân tích ngôn ngữ tự nhiên để tạo ra câu truy vấn SPARQL (Vijayarajan et al., 2016), xây dựng công cụ I2T (*Image to Text*) nhằm tạo ra các RDF mô tả ngữ nghĩa hình ảnh (Yao et al., 2010)... Theo kết quả các công trình đã công bố gần đây cho thấy phương pháp tìm kiếm ảnh bằng kỹ thuật gom cụm và ontology là khả thi và đạt kết quả tốt.

Vijayarajan và cộng sự (2016) thực hiện tìm kiếm ảnh dựa trên câu truy vấn SPARQL bằng ngôn ngữ RDF (Vijayarajan et al., 2016) bằng cách dựa trên mô tả nội dung hình ảnh. Phương pháp này chưa thực hiện phân lớp hình ảnh từ các đặc trưng màu sắc và đặc trưng không gian để tạo các từ khóa trước khi thực hiện tra cứu. Trong khi đó, Yue Cao và cộng sự (2016) đã sử dụng mạng CNN (*Convolutional Neural Network*) phân lớp hình ảnh để tạo ra các vector đặc trưng nhị phân. Trên cơ sở này, nhóm tác giả đã đề xuất mô hình DVSH (*Deep Visual-Semantic Hashing*) nhằm xác định tập ảnh tương tự với ảnh đầu vào (Cao et al., 2016). Kết quả thực nghiệm đề xuất đã minh chứng tính hiệu quả của phương pháp đề xuất trong công trình này và có thể cải tiến cho những công trình tiếp theo.

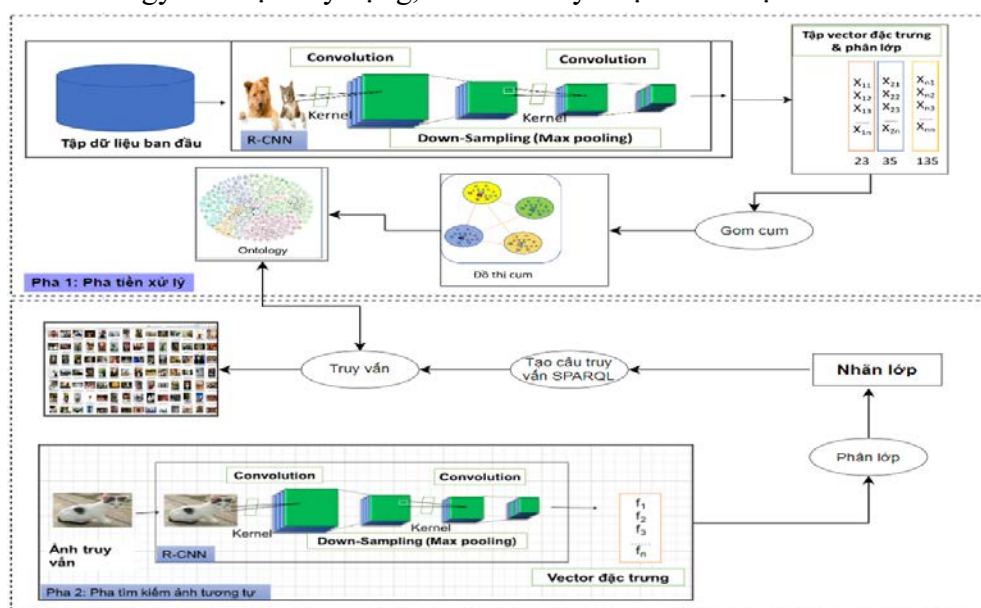
Sau đó, M. N. Asim và cộng sự (2019), đã xem xét các phương pháp truy xuất thông tin dựa trên ontology áp dụng cho truy vấn văn bản và dữ liệu đa phương tiện. Nhóm tác giả đã so sánh hiệu suất với các phương pháp tiếp cận trước đó về truy vấn dữ liệu đa phương

tiện cao hơn một số công trình cùng lĩnh vực trước đó. Trong công trình này, tác giả sử dụng ngôn ngữ bộ ba RDF để thực hiện lưu trữ và truy vấn trên ontology (Asim et al., 2019). Tuy nhiên, nhóm tác giả chưa đề cập đến kết quả thực nghiệm cụ thể để so sánh với các công trình trước đây nhằm minh chứng tính khả thi và hiệu quả của phương pháp đề xuất, đây cũng là một thiếu sót cần được nêu rõ từ những cải tiến sau này.

2. Đối tượng và phương pháp nghiên cứu

2.1. Mô hình truy vấn ảnh đề xuất

Để đánh giá tính đúng đắn của phương pháp đề xuất áp dụng cho bài toán tìm kiếm ảnh, một mô hình truy vấn ảnh được đề xuất dựa trên mạng nơ-ron tích chập kết hợp với đồ thị cụm và ontology đã được xây dựng, mô hình này được minh họa như Hình 1.



Hình 1. Mô hình truy vấn ảnh dựa trên mạng nơ-ron tích chập và đồ thị cụm

Mô hình tìm kiếm ảnh dựa trên mạng nơ-ron tích chập và đồ thị cụm được chia thành hai pha: pha tiền xử lý và pha truy vấn, cụ thể các bước như sau:

Pha tiền xử lý

Bước 1. Tạo đồ thị cụm từ tập vec-tơ đặc trưng thị giác của từng phân vùng hình ảnh;

Bước 2. Đánh giá độ tương tự giữa các đỉnh của đồ thị;

Bước 3. Tạo Ontology lưu trữ cho từng nhóm hình ảnh trên đồ thị phân cụm.

Pha tìm kiếm ảnh tương tự

Bước 1. Trích xuất đặc trưng thị giác và phân lớp hình ảnh;

Bước 2. Tạo câu truy vấn SPARQL dựa trên các phân lớp của hình ảnh đầu vào và các nhóm đối tượng láng giềng;

Bước 3. Kết xuất các hình ảnh tương tự và sắp xếp theo độ đo tương tự với ảnh truy vấn.

2.2. Xây dựng ontology cho bài toán tìm kiếm ảnh

Để xây dựng ontology đáp ứng yêu cầu cho mô hình truy vấn ảnh đã đề xuất; một số cơ sở lý thuyết được trình bày gồm: mạng nơ-ron tích chập và đồ thị cụm cụ thể như sau:

2.2.1. Mạng nơ-ron tích chập

Trong bài báo này, lí thuyết về mạng tích chập được tiếp cận để ứng dụng phân lớp hình ảnh và trích xuất các đặc trưng làm đầu vào cho quá trình xây dựng đồ thị cụm. Từ đó, làm cơ sở để xây dựng ontology đáp ứng mô hình tìm kiếm ảnh đã đề xuất. Cấu trúc mạng nơ-ron tích chập được ứng dụng để kiểm chứng cho quá trình nhận diện và phân lớp các đối tượng trên ảnh bộ ảnh MS-COCO và Flickr. Cấu trúc mạng R-CNN được trình bày về lí thuyết và minh họa ứng dụng. Cấu trúc mạng Faster R-CNN được minh họa như Hình 2, gồm các thành phần: Region Proposal Network (RPN); Lớp RoI pooling; Detection network (Bharati & Pramanik, 2020).

2.2.2. Xây dựng đồ thị cụm

Mỗi hình ảnh được trích xuất các đối tượng và tạo ra các ảnh đối tượng, mỗi ảnh đối tượng được trích xuất các đặc trưng để làm cơ sở gom nhóm các hình ảnh trên từng cụm theo thuật toán K-means. Mỗi đỉnh của đồ thị láng giềng là một cụm gồm các hình ảnh có thành phần đặc trưng tương tự theo độ đo khoảng cách Euclid. Sau khi phân cụm, các vector đại diện của cụm được tính bằng cách lấy giá trị trung bình để từ đó đánh giá khoảng cách giữa các đỉnh cụm. Nếu các đỉnh cụm này nhỏ hơn một ngưỡng θ cho trước thì hai đỉnh đó được gọi là láng giềng và có trọng số là khoảng cách Euclid được chuẩn hóa trên đoạn $[0,1]$.

Tập dữ liệu ảnh được trích xuất các hình ảnh con và nhãn tương ứng dựa trên mạng nơ-ron tích chập, tập các phần tử này được phân bố dựa trên thuật toán gom cụm k-means với số cụm được tăng trưởng dần theo độ đo khoảng cách Euclid và theo hai ngưỡng ϵ , θ . Mỗi phần tử lưu trữ đặc trưng các hình ảnh con và nhãn lớp là một Element gồm có các thành phần: định danh ảnh gốc (*ImagesID*), định danh ảnh thành phần (*ObjectID*), vec-tơ đặc trưng f của hình ảnh, nhãn lớp (*ObjLabel*). Mỗi thành phần được kí hiệu là:

$$Element = \langle ImagesID, ObjectID, f, ObjLabel \rangle \quad (1)$$

Mỗi một đỉnh của đồ thị là một tập các phần tử *Element* có độ tương tự dựa trên độ đo Euclid và giữa các phần tử này có khoảng cách nhỏ hơn ngưỡng ϵ cho trước. Trong mỗi đỉnh này thực hiện quá trình đánh dấu láng giềng theo một ngưỡng trong một cụm. Tập các phần tử trong một cụm được kí hiệu là:

$$C = \{Element_i | i = 1..M\} \quad (2)$$

Trong đó M là số phần tử *Element* của một đỉnh cụm. Các phần tử trong một đỉnh cụm có một nhãn lớp, nên nhãn lớp đại diện cho đỉnh cụm đó là nhãn lớp xuất hiện nhiều nhất, do đó các phần tử thuộc một đỉnh cụm, thỏa điều kiện độ đo giữa các đặc trưng của các *Element* nhỏ hơn một khoảng cách ϵ cho trước.

Thuật toán phân bố các phần tử vào trong các cụm của đồ thị được thực hiện bằng cách phân bố từng hình ảnh vào trong các cụm đã có ở bước trước với đầu vào là một tập các *Element*, đầu ra là một tập các đỉnh cụm và nhãn tương ứng.

Thuật toán 1: Gom cụm các đỉnh cho đồ thị

Đầu vào: Tập $J = \{Element_i \mid i = 1..N\}$ của tập ảnh ban đầu, ngưỡng ε .

Đầu ra: Tập các cụm $C = \{C_i \mid i = 1..k\}$

Begin

If $C = \emptyset$ **then**

$C_i = C_1 \cup \{Element_1\};$

$C = C \cup \{C_1\}; k = 1;$

EndIf

Foreach ($Element_i \in J$) **do**

$m = \operatorname{argmin} \{Euclid(f_i, C_j) \mid j = 1..k\};$

If ($Euclid(f_i, C_m) > \varepsilon$) **then**

$k = k + 1;$

$C_k = C_k \cup \{Element_i\};$

else

$C_m = C_m \cup \{Element_i\};$

EndIf;

EndForeach

return C;

End.

Sau khi thực hiện tạo các cụm, các nhãn được gán cho mỗi cụm theo **thuật toán 2**.

Thuật toán 2: Gán nhãn cho đỉnh cụm

Đầu vào: Tập các cụm $C = \{C_i \mid i = 1..k\}$, tập các nhãn $L = \{label_i \mid i = 1..l\}$

Đầu ra: Tập các cụm $CL = \{cl_i \mid i = 1..k\}$

Begin

Khởi tạo các nhãn cụm

For $i = 1$ **to** k **do**

$cl_i = null$

Endfor

Gán nhãn đỉnh cụm

For $i = 1$ **to** k **do**

For $j = 1$ **to** $|C_i|$ **do**

$\text{Count}[Element_j, label] = \text{Count}[Element_j, label] + 1;$

Endfor

$m = \operatorname{argMax} \{\text{Count}[i] \mid i = 1..l\};$

$cl_i = label_m;$

Endfor

Return CL;

End.

Mỗi đỉnh của đồ thị được tính trung bình và tạo ra phần tử đại diện cho cụm đó, từ đó khoảng cách hai cụm được tính bằng độ đo Euclid và quy chuẩn trên đoạn $[0,1]$. **Thuật toán 3** đánh trọng số cho đồ thị.

Thuật toán 3: Tạo cung cho đồ thị cụm

Đầu vào: Tập các cụm $C = \{C_i \mid i = 1..k\}$, ngưỡng θ

Đầu ra: Ma trận trọng số M giữa các đỉnh cụm.

Begin

 Khởi tạo trọng số đồ thị.

For $i = 1$ **to** k **do**

$M[s,t] = \infty$ với $s,t = 1..k; s \neq t$

$M[i,i] = 0;$

EndFor

 Tạo trọng số cho đồ thị.

For $i = 1$ **to** $k - 1$ **do**

For $j = i + 1$ **to** k **do**

If $(\text{Euclid}(C_i, C_j) < \theta)$ **then**

$M[i,j] = \text{Euclid}(C_i, C_j); M[j,i] = \text{Euclid}(C_i, C_j);$

EndIf

EndFor

EndFor

Return M;

End.

2.2.3. Xây dựng ontology

Với mỗi đỉnh cụm trên đồ thị, tập các ảnh thành phần được trích xuất và đưa vào ontology theo phân lớp đại diện. Các hình ảnh gốc được đưa vào tương ứng với các ảnh đối tượng thuộc ontology. **Thuật toán 4** tích hợp hình ảnh vào ontology được thực hiện như sau:

Thuật toán 4: Tích hợp hình ảnh vào ontology

Đầu vào: Đồ thị cụm $G = (V, E)$, ontology O)

Đầu ra: ontology O đã được làm giàu.

Begin

Foreach $(V_i \in G)$ **do**

For $j = 1$ **to** $|V_i|$ **do**

 Insert $Element_i$ into ontology O;

 Insert Image into ontology O;

EndFor

EndForeach

Return ontology O;

End.

3. Kết quả và thảo luận

3.1. Môi trường và dữ liệu thực nghiệm

Hệ truy vấn ảnh được thực nghiệm trên cơ sở hai giai đoạn gồm tiền xử lí ảnh và truy vấn ảnh; đối với giai đoạn tiền xử lí ảnh, hệ thống được thực thi trên hệ thống máy Server với cấu hình CPU Xeon(R) Gold 6258R CPU 2.70Ghz, Ổ cứng SSD 1024GB, Ram 16GB, hệ điều hành Server Datacenter 2019. Xây dựng đồ thị phân cụm và ontology được thực hiện trên máy Server như trên với hệ thống chương trình được thực thi trên ngôn ngữ lập trình C#, .Net Frameworks 4.8. Quá trình tìm kiếm ảnh của người dùng được với cấu hình máy tính: Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz 2.30 GHz Installed RAM 8.00 GB và hệ điều hành Windows 10 Professional.

Bảng 1. Mô tả phân chia bộ ảnh MS-COCO và Flickr trong thực nghiệm

Tập ảnh	Số ảnh	Số ảnh tạo đồ thị	Số ảnh Testing	Số ảnh Validation
MS-COCO	163,957	118,287	40,670	5,000
Flickr	31,783	29,000	1,783	1,000

3.2. Kết quả thực nghiệm và đánh giá

Để thực hiện truy vấn tập ảnh tương tự với ảnh đầu vào, **thuật toán 5** thực hiện truy vấn ảnh dựa trên ontology như sau:

Thuật toán 5: Thuật toán truy vấn ảnh.

Đầu vào: Ảnh truy vấn I

Đầu ra: Tập ảnh tương tự SI ;

Begin

Khởi tạo tập nhãn lớp: $L = \emptyset$;

Phân lớp ảnh đầu vào I để có tập nhãn L_p ;

$L = L \cup L_p$;

Foreach ($l_i \in L$) **do**

$L_g = \{l_g \in G \mid l_g \text{ là láng giềng của } l_i\}$; $L = L \cup L_g$;

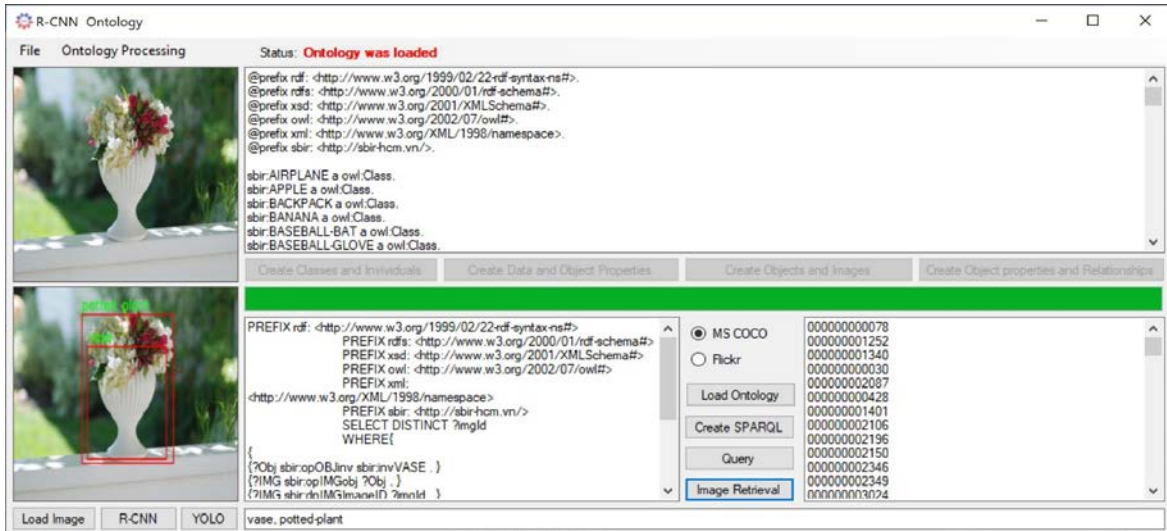
EndForeach

Tạo câu truy vấn SPARQL bằng thuật toán 3.5.

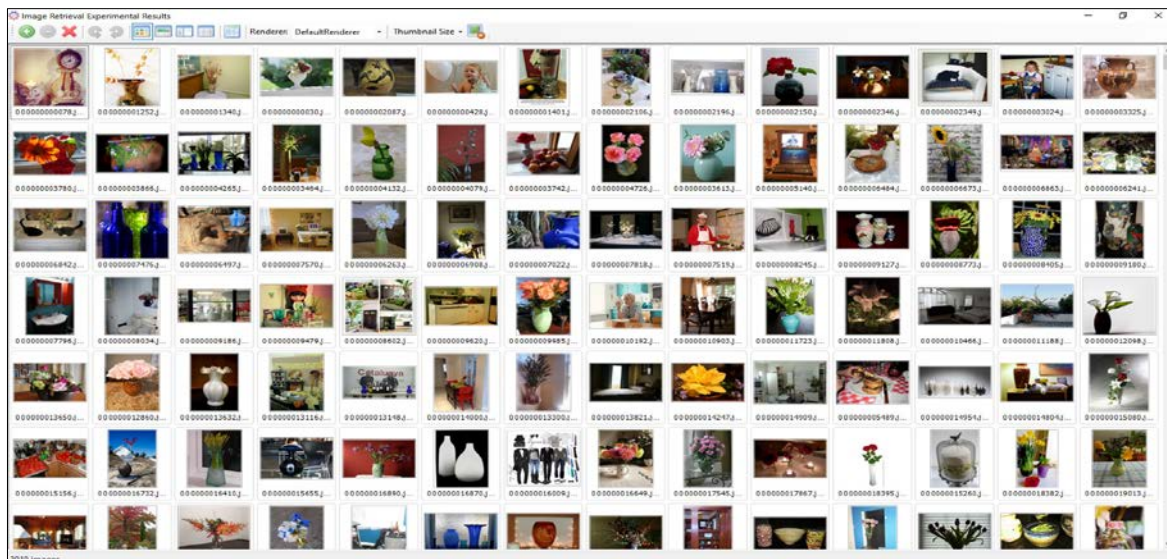
$SI = \text{Query on ontology } O$;

Return SI ;

End



Hình 4. Hệ truy vấn ảnh dựa trên mạng neuron và đồ thị cụm (SBIR)



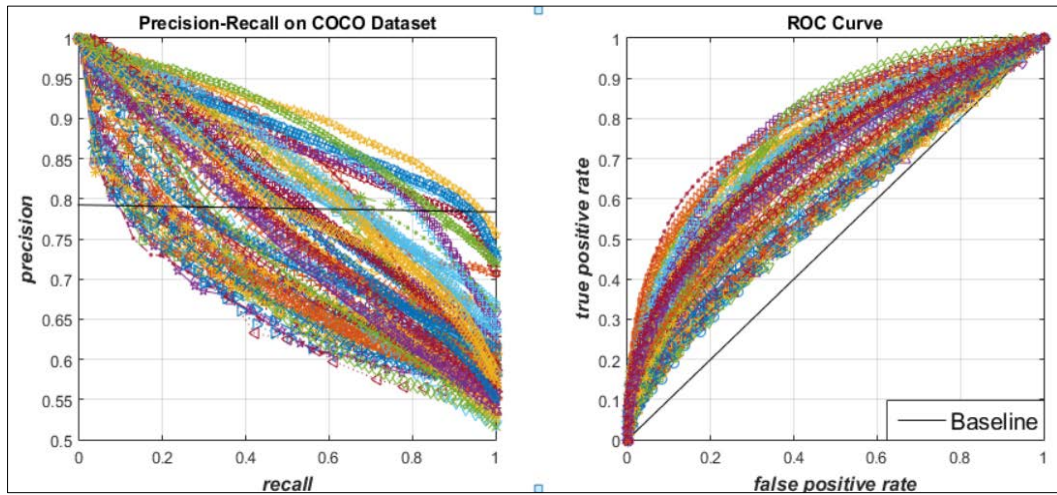
Hình 5. Một kết quả truy vấn tập ảnh tương tự trên bộ ảnh MS-COCO

Sau khi truyền một ảnh đầu vào, quá trình truy vấn tập ảnh tương tự được minh họa như Hình 4 gồm: (1) chọn Ontology của bộ ảnh MS-COCO để truy vấn, với ảnh đầu vào thuộc bộ MS-COCO có mã số 00000000030.jpg, ảnh này được phân thành hai lớp gồm *vase* và *potted-plant*, hai nhãn lớp này được làm cơ sở để xây dựng câu truy vấn SPARQL nhằm truy vấn và truy hồi hình ảnh kết quả. Sau khi thực hiện truy vấn, tập ảnh tương tự với ảnh đầu vào được minh họa như Hình 5.

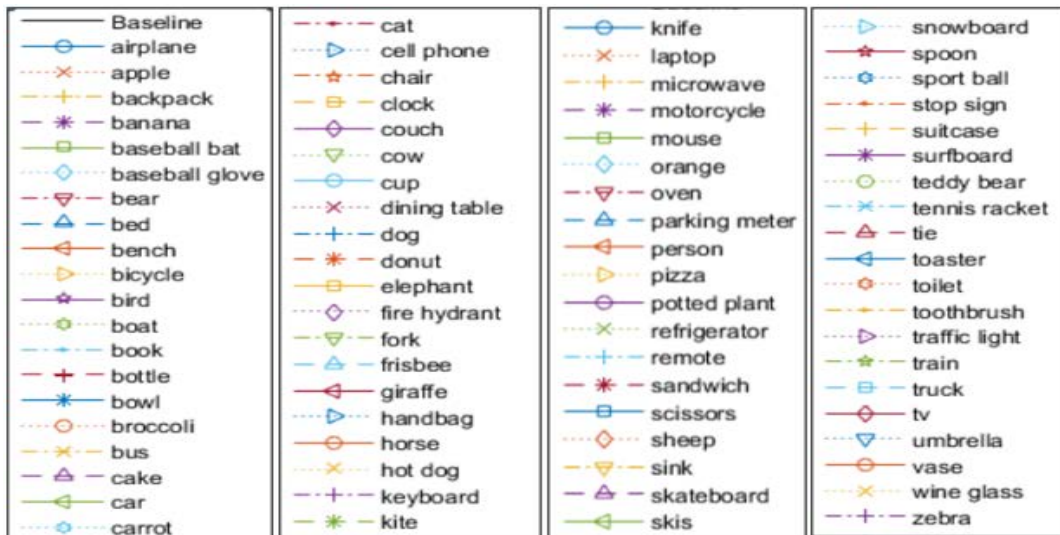
Bảng 2. Hiệu suất tìm kiếm ảnh của phương pháp đề xuất trên các bộ dữ liệu

Tập ảnh	Precision	Recall	F-measure	Query Time (MS-COCO)
MS-COCO	0.7950	0.6437	0.7114	78.77
Flickr	0.8116	0.6819	0.7411	68.83

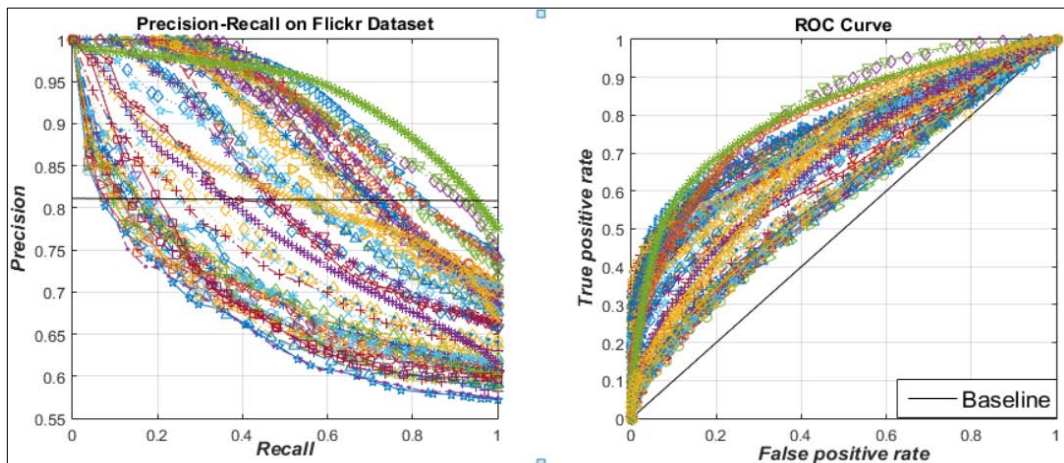
Các đồ thị Precision, Recall và đường cong ROC bao gồm nhiều đường cong, mỗi đường cong mô tả độ chính xác (*precision*) và độ phủ (*recall*) của một chủ đề trong bộ dữ liệu MS-COCO, Flickr. Đường cong trong đồ thị ROC cho biết tỉ lệ truy vấn đúng và sai. Hình 6 và 8 mô tả hiệu suất và tính đúng đắn của kết quả truy vấn trên các bộ ảnh COCO, Flickr. Đồ thị cho thấy tính chính xác của hệ truy vấn tập ảnh COCO nằm tập trung ở vùng [0.52, 1.0]; độ chính xác của tập ảnh Flickr nằm tập trung ở vùng [0.48, 1.0].



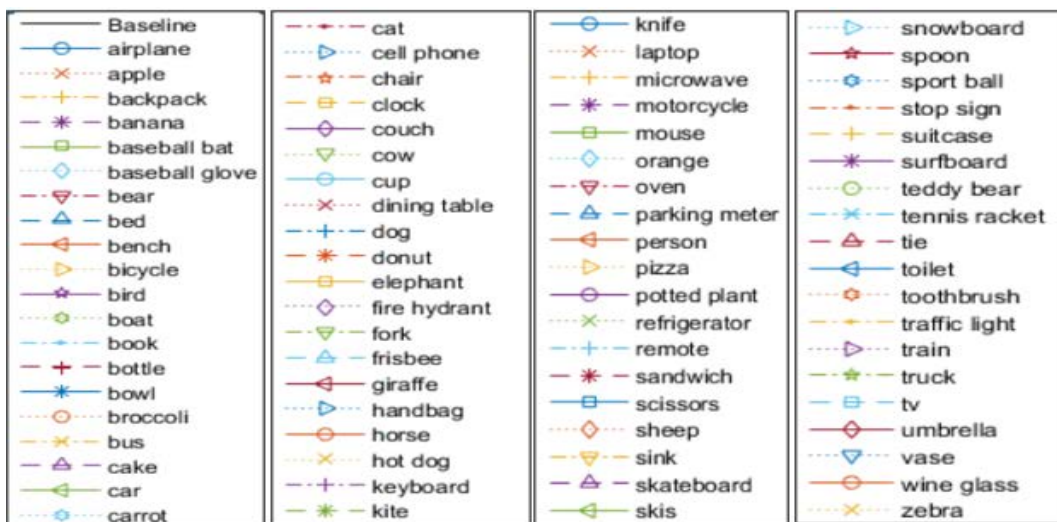
Hình 6. Precision-Recall và đường cong ROC bộ ảnh MS-COCO



Hình 7. Các chú thích cho đồ thị tại Hình 6

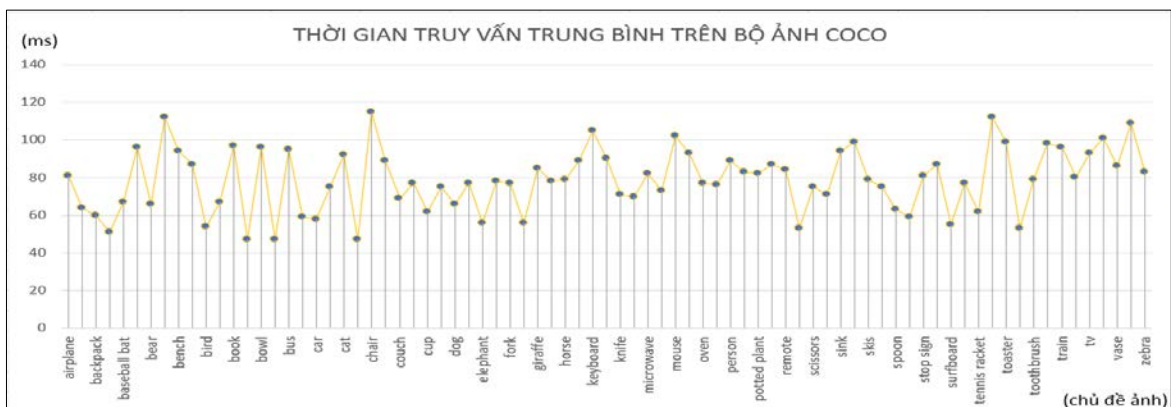


Hình 8. Precision-Recall và đường cong ROC bộ ảnh Flickr

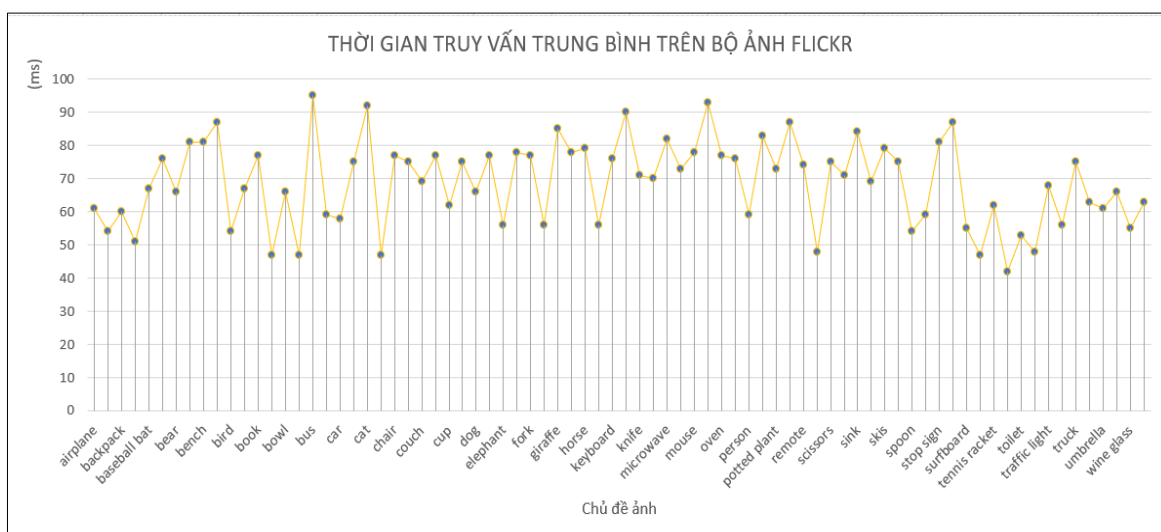


Hình 9. Chú thích cho đồ thị tại Hình 8

Biểu đồ biểu diễn thời gian truy vấn trung bình trên các bộ ảnh MS-COCO, Flickr được minh họa trong Hình 10 và 11. Trong đó mỗi thư mục ảnh được biểu diễn bằng một đường trung bình cho thời gian truy vấn.



Hình 10. Thời gian truy vấn trung bình theo chủ đề của bộ ảnh MS-COCO



Hình 11. Thời gian truy vấn trung bình theo chủ đề của bộ ảnh Flickr

Để minh chứng tính hiệu quả của mô hình truy vấn ảnh đã đề xuất; các kết quả này được so sánh với một số công trình cùng lĩnh vực trong nhữn năm gần đây được thể hiện trong Bảng 3, 4. Điều này cho thấy việc kết hợp mạng R-CNN với đồ thị cụm để thực hiện tìm kiếm ảnh trên ontology là khả thi và hiệu quả.

Bảng 3. So sánh hiệu suất truy vấn giữa các phương pháp trên bộ dữ liệu MS-COCO

Phương pháp	Bộ dữ liệu	Độ chính xác trung bình
CNN - RNN, 2016 (Song et al., 2018)	COCO	0.6120
CAM, TopK = 5, 2019 (Wang et al., 2019)	COCO	0.6890
SBIR	COCO	0.7950

Bảng 4. So sánh hiệu suất truy vấn giữa các phương pháp trên bộ dữ liệu Flickr

Phương pháp	Bộ dữ liệu	Độ chính xác trung bình
BGAN – 48 bit, 2018 (Song et al., 2018)	Flickr	0.7030
CAM, TopK = 5, 2019 (Wang et al., 2019)	Flickr	0.7710
SBIR	Flickr	0.8116

Theo kết quả thực nghiệm cho thấy, hệ truy vấn ảnh **SBIR** với hiệu suất truy vấn các hơn các công trình cùng bộ dữ liệu là bởi các lí do sau:

- 1) Hệ truy vấn ảnh **SBIR** kết hợp được các kĩ thuật học máy R-CNN, YOLO;
- 2) Hệ truy vấn ảnh **SBIR** thực hiện phân lớp ảnh trước khi tìm kiếm ảnh;
- 3) Kết hợp cấu trúc dữ liệu đồ thị cụm và ontology đã nâng cao hiệu suất tìm kiếm ảnh.

4. Kết luận và kiến nghị

Trong bài báo một mô hình tìm kiếm ảnh sử dụng mạng nơ-ron tích chập kết hợp cấu trúc đồ thị cụm nhằm nâng cao hiệu suất truy vấn ảnh và giảm thời gian tìm kiếm trên các bộ ảnh đa đối tượng MS-COCO, Flickr. Một số thuật toán đề xuất bao gồm: gom cụm các đỉnh cho đồ thị cụm; gán nhãn cho đỉnh cụm; tạo cung đồ thị cụm; tích hợp dữ liệu vào ontology và thuật toán truy vấn ảnh trên ontology. Kết quả thực nghiệm về hiệu suất truy vấn ảnh tương ứng trên bộ ảnh MS-COCO, Flickr tương ứng là **0.7950, 0.8116**; đồng thời so sánh và phân tích lí do tại sao kết quả này cao hơn một số công trình khác đã công bố trên cùng bộ ảnh thực nghiệm.

- ❖ **Tuyên bố về quyền lợi:** Các tác giả xác nhận hoàn toàn không có xung đột về quyền lợi.
- ❖ **Lời cảm ơn:** Chúng tôi xin trân trọng cảm ơn Trường Đại học Sư phạm Thành phố Hồ Chí Minh, Khoa Công nghệ Thông tin – Trường Đại học Khoa học, Đại học Huế đã góp ý chuyên môn cho nghiên cứu; Trường THCS Bình Tân – Lagi - Bình Thuận, Trường Đại học Công nghiệp Thực phẩm Thành phố Hồ Chí Minh, Bảo hiểm xã hội Tỉnh Bà Rịa – Vũng Tàu đã tạo điều kiện về cơ sở vật chất giúp chúng tôi hoàn thành bài nghiên cứu này.

TÀI LIỆU THAM KHẢO

- Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. (2019). The use of ontology in retrieval: a study on textual, multilingual, and multimedia retrieval. *IEEE Access*, 7, 21662-21686.
- Bharati, P., & Pramanik, A. (2020). Deep learning techniques—R-CNN to mask R-CNN: A survey. *Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019*, 657-668.
- Cao, Y., Long, M., Wang, J., Yang, Q., & Yu, P. S. (2016). *Deep visual-semantic hashing for cross-modal retrieval*. Paper presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Cevikalp, H., Elmas, M., & Ozkan, S. (2018). Large-scale image retrieval using transductive support vector machines. *Computer Vision and Image Understanding*, 173, 2-12.
- Dinh, N. T., Le, T. M., & Van, T. T. (2022). An Improvement Method of Kd-Tree Using k-Means and k-NN for Semantic-Based Image Retrieval System *Information Systems and Technologies: WorldCIST 2022, Volume 2* (pp. 177-187). Springer.
- Dinh, N. T., Nhi, N. T. U., Le, T. M., & Van, T. T. (2023). A model of image retrieval based on KD-Tree Random Forest. *Data Technologies and Applications*.
- Flickr. (2017). Dataset Flickr 2017. Retrieved from <https://www.kaggle.com/datasets/hsankesara/flickr-image-dataset>

- Jabeen, S., Mehmood, Z., Mahmood, T., Saba, T., Rehman, A., & Mahmood, M. T. (2018). An effective content-based image retrieval technique for image visuals representation based on the bag-of-visual-words model. *PloS one*, *13*(4), e0194526.
- Kumar, A., Dyer, S., Kim, J., Li, C., Leong, P. H., Fulham, M., & Feng, D. (2016). Adapting content-based image retrieval techniques for the semantic annotation of medical images. *Computerized Medical Imaging and Graphics*, *49*, 37-45. DOI:10.1016/j.compmedimag.2016.01.001
- Lin, C. H., Chen, C. C., Lee, H. L., & Liao, J. R. (2014). Fast K-means algorithm based on a level histogram for image retrieval. *Expert Systems with Applications*, *41*(7), 3276-3283.
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). *Microsoft coco: Common objects in context*. Paper presented at the Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13.
- MS-COCO. (2017). Dataset MS-COCO 2017. Retrieved from <https://www.kaggle.com/datasets/awsaf49/coco-2017-dataset?resource=download>
- Ptucha, R., Such, F. P., Pillai, S., Brockler, F., Singh, V., & Hutkowsky, P. (2019). Intelligent character recognition using fully convolutional neural networks. *Pattern Recognition*, *88*, 604-613.
- Saboorian, M. M., Jamzad, M., & Rabiee, H. R. (2010). *User adaptive clustering for large image databases*. Paper presented at the 2010 20th International Conference on Pattern Recognition.
- Song, J., He, T., Gao, L., Xu, X., Hanjalic, A., & Shen, H. T. (2018). Binary generative adversarial networks for image retrieval. Paper presented at the *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Spanier, A. B., & Joskowicz, D. C., L. (2017). A new method for the automatic retrieval of medical cases based on the RadLex ontology. *International journal of computer assisted radiology and surgery*, *12*(3), 471-484.
- Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. *Human-centric Computing and Information Sciences*, *6*(1), 1-30.
- Wang, Z., Liu, X., Li, H., Sheng, L., Yan, J., Wang, X., & Shao, J. (2019). *Camp: Cross-modal adaptive message passing for text-image retrieval*. Paper presented at the *Proceedings of the IEEE/CVF international conference on computer vision*.
- Yao, B. Z., Yang, X., Lin, L., Lee, M. W., & Zhu, S. C. (2010). I2t: Image parsing to text description. *Proceedings of the IEEE*, *98*(8), 1485-1508.

**IMAGE RETRIEVAL USING CONVOLUTIONAL NEURAL NETWORKS
AND CLUSTER GRAPH****Pham Hoang Phuong¹, Do Xuan Hiep², Nguyen Thi Dinh³, Van The Thanh^{4*}**¹*Binh Tan High School, Lagi, Binh Thuan, Vietnam*²*Social Insurance in Ba Ria – Vung Tau Province, Vietnam*³*HCMC University of Industry and Trade, Vietnam*⁴*Ho Chi Minh City University of Education, Vietnam***Corresponding author: Van The Thanh – Email: thanhvt@hcmue.edu.vn**Received: October 11, 2022; Revised: November 05, 2022; Accepted: June 26, 2023***ABSTRACT**

In this paper, a model of image retrieval using a convolutional neural network combined with a cluster graph is implemented to improve performance and reduce image query time. To implement this problem: (1) a convolutional neural network was used to identify and classify objects on the image; (2) a cluster graph structure was built to perform ontology construction; (3) similar image sets were extracted based on the following ontology performed when searching by SPARQL query. For each input image, after classifying each object using a convolutional neural network and feature vector extraction, it was classified followed by being retrieved on ontology to extract a set of similar images. Based on the proposed theory, a model of image retrieval is proposed and experimented on COCO and Flickr images datasets with the corresponding accuracy of 0.7950 and 0.8116, respectively. According to the results, the proposed method is evaluated as correct based on the comparison with other works on the same set of images. The proposed model also works to different data sets.

Keywords: convolutional neural networks; image retrieval; similar images; SPARQL