Tạp chí Khoa học Trường Đại học Cần Thơ
Phần A: Khoa học Tự nhiên, Công nghệ và Môi trường
website: sj.ctu.edu.vn

DOI:10.22144/ctu.jsi.2017.008

TÌM KIẾM ẢNH THEO NỘI DUNG VÀ NGỮ NGHĨA

Lư Minh Phúc và Trần Công Án
Khoa Công nghệ Thông tin và Truyền thông, Trường Đại học Cần Thơ

Thông tin chung:
Ngày nhận bài: 15/09/2017
Ngày nhận bài sửa: 10/10/2017
Ngày duyệt đăng: 20/10/2017

Title:
Semantic content-based image search

Từ khóa:
Học sâu, mạng neural tích chập, ngữ nghĩa, ontology, SPARQL, tìm kiếm ảnh

Keywords:
Convolutional neural network, deep learning, image search, semantics, ontology, SPARQL

ABSTRACT

Content-based image search has been concerned recently. This search method helps to overcome shortcomings of current meta-data-based search method, which is sensitive to the meta-data enclosed with images. In this paper, a content-based image search system is developed based on the convolutional neural network deep learning model. In addition, the search system is also combined with semenatic search technique that enables the improvement of the search result. The semance searching capacity bases on a domain-ontology that describes semantic relationships among image topics. The experimental result shows that the accuracy of the convolutional neural network classification model on the test set is 85.75%. Moreover, the semantic search is helpful to widen and improve the search result significantly, particularly in the case that the searching keywords is ambigous or unclear.

TÓM TẮT

Trong những năm gần đây, các nghiên cứu về tìm kiếm ảnh theo nội dung đang được quan tâm vì phương pháp tìm kiếm này có thể khắc phục nhược điểm của phương pháp tìm kiếm dựa trên meta data là không bị ảnh hưởng bởi sự thiếu hoặc sai của meta data kèm theo ảnh. Trong nghiên cứu này, chúng tôi sẽ xây dựng một hệ thống tìm kiếm ảnh theo nội dung dựa trên việc phân loại tập ảnh theo nội dung bằng mô hình mạng neural tích chập (CNNs) của kỹ thuật học sâu (deep learning). Đồng thời, chúng tôi sẽ kết hợp ngữ nghĩa vào quá trình tìm kiếm để cho phép mở rộng thêm kết quả tìm kiếm ảnh theo những khái niệm ngữ nghĩa mà con người đã chấp nhận, so với ý nghĩa của những thông tin có được từ những đặc trưng của ảnh. Việc kết hợp ngữ nghĩa vào quá trình tìm kiếm sẽ dựa trên một domain ontology do chúng tôi xây dựng để mô tả các mối quan hệ ngữ nghĩa giữa các chủ đề ảnh. Kết quả thực nghiệm cho thấy mô hình CNNs phân lớp tập ảnh kiểm thử đạt độ chính xác là 85,75% và việc kết hợp ngữ nghĩa cho phép mở rộng và đa dạng hóa kết quả tìm kiếm, đặc biệt hữu ích trong các trường hợp từ khóa tìm kiếm có nhiều từ đồng nghĩa hoặc nhập nhằng.

Trích dẫn: Lư Minh Phúc và Trần Công Án, 2017. Tìm kiếm ảnh theo nội dung và ngữ nghĩa. Tạp chí Khoa học Trường Đại học Cần Thơ. Số chuyên đề: Công nghệ thông tin: 58-64.

1 GIỚI THIỆU

Cùng với sự phát triển vượt trội của các công nghệ kỹ thuật số và sự phổ biến rộng rãi các thiết bị quay phim, chụp ảnh dẫn đến kho dữ liệu ảnh lưu trữ trên Web cũng tăng theo một cách nhanh chóng. Mary Meeker, một chuyên gia về phân tích Internet và công nghệ thuộc đại học Cornell (Mỹ) trong báo cáo thường niên về xu hướng Internet cho biết: "*Chúng ta đã tải lên mạng trung bình khoảng 1,8 tỷ ảnh số trong một ngày và 657 tỷ bức ảnh trong một năm. Có nghĩa là cứ mỗi hai phút thì số lượng ảnh*

chúng ta chụp sẽ nhiều hơn tổng số ảnh đã có của 150 năm về trước" (Meeker, 2014). Đây là một thách thức lớn cho việc tổ chức và tìm kiếm ảnh theo cách truyền thống. Vì vậy, việc xây dựng một hệ thống tìm kiếm ảnh là một điều cấp bách và cần thiết. Các hệ thống tìm kiếm ảnh hiện tại thường sử dụng phương pháp là tìm kiếm ảnh theo các văn bản đi kèm với ảnh (meta-data) hoặc theo nội dung (sự tương đồng) của ảnh giúp cho việc tìm kiếm đơn giản và hiệu quả. Tuy nhiên, hai phương pháp tìm kiếm ảnh trên vẫn còn một số hạn chế làm cho kết quả tìm kiếm chưa chính xác hoặc chưa làm hài lòng hoàn toàn người sử dụng. Đối với phương pháp tìm kiếm ảnh dựa trên văn bản hoặc các mô tả (meta-data) kèm theo ảnh sẽ không chính xác khi các mô tả này bị sai sót hoặc không tồn tại.

Kế đến là phương pháp tìm kiếm ảnh theo nội dung "truyền thống" thường dựa vào các đặc trưng trực quan như màu sắc, kết cấu, hình dạng, đặc trưng cục bộ được rút trích từ ảnh. Phương pháp này có hạn chế là làm cách nào để xác định và chọn ra được những đặc trưng đại diện có ảnh hưởng cao đến độ chính xác của kết quả tìm kiếm? Quá trình chọn lựa này sẽ gây mất nhiều thời gian trong quá trình xây dựng hệ thống; ngoài ra, còn phát sinh vấn đề do sự cách biệt ngữ nghĩa (semantic gap) giữa đặc trưng ở mức thấp dưới dạng các pixel ảnh và mức khái niệm cao theo sự chấp nhận của con người như sunset, dog,...

Do đó, trong nghiên cứu này, chúng tôi sẽ đề xuất một phương pháp để xây dựng một hệ thống tìm kiếm ảnh theo nội dung dựa trên mô hình học sâu là mạng neural tích chập (CNNs) nhằm tận dụng tối đa sức mạnh tính toán của máy tính trong việc tìm kiếm hình ảnh theo nội dung. Đồng thời, hệ thống cũng tích hợp ngữ nghĩa vào việc tìm kiếm dựa trên một domain-ontology để mô tả các mối quan hệ giữa các chủ đề ảnh cần phân lớp. Phương pháp tìm kiếm này không những khắc phục được các hạn chế của phương pháp tìm kiếm dựa trên meta-data mà còn cho phép mở rộng và đa dạng hóa kết quả tìm kiếm thông qua việc kết hợp ngữ nghĩa vào việc tìm kiếm.

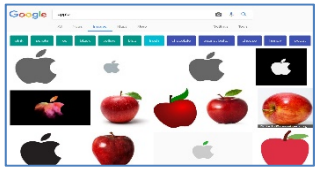
Bài báo này bao gồm 5 phần. Phần một giới thiệu sự cần thiết của các hệ thống tìm kiếm ảnh và nhược điểm của phương pháp tìm kiếm ảnh theo meta-data. Phần hai tóm lược các nghiên cứu có liên quan. Phần ba trình bày về kiến trúc của hệ thống, phân loại ảnh bằng CNN và tìm kiếm theo ngữ nghĩa. Phần bốn đánh giá kết quả tìm kiếm ảnh qua thực nghiệm và phần cuối sẽ trình bày kết luận về nghiên cứu.

2 CÁC NGHIÊN CỨU CÓ LIÊN QUAN

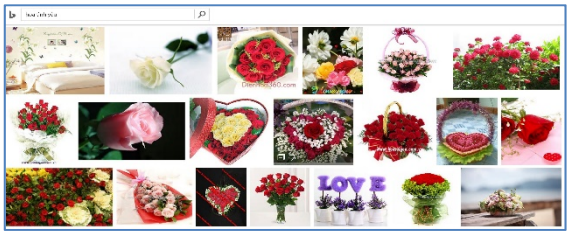
Hiện nay, có nhiều công cụ và công trình nghiên cứu khác nhau liên quan đến việc xây dựng hệ thống tìm kiếm ảnh nhằm cải tiến hiệu quả tìm kiếm ảnh để đáp ứng nhu cầu của người dùng ngày càng tốt hơn.

Google Images Search là một trong các công cụ tìm kiếm ảnh được sử dụng phổ biến nhất hiện nay. Công cụ này cho phép người sử dụng nhập các từ khóa liên quan đến ảnh cần tìm và thực hiện việc tìm kiếm thông qua việc phân tích các meta-data và văn bản đi kèm với ảnh. Phương pháp này cho kết quả tương đối tốt, đáp ứng nhu cầu cơ bản của người sử dụng. Tuy nhiên, các kết quả trả về sẽ không đúng với yêu cầu đặt ra khi các meta-data đi kèm với ảnh bị thiếu hoặc sai sót và khi những từ khóa truy vấn mang ý nghĩa nhập nhằng. Ví dụ, với truy vấn "*apple*" để tìm hình quả táo thì kết quả trả về đầu tiên không thỏa mãn như được minh họa trong Hình 1.

Bing cũng là một trong các bộ máy tìm kiếm thông dụng được phát triển bởi Microsoft. Đây là một bộ máy tìm kiếm ảnh mạnh mẽ với cơ sở dữ liệu ảnh lớn. Bing cho phép người dùng tìm kiếm ảnh bằng cách nhập câu truy vấn ảnh và tìm kiếm dựa trên các meta-data hoặc văn bản đi kèm với ảnh. Cũng tương tự như Google Images Search, công cụ tìm kiếm này cũng gặp những vấn đề đã đề cập bên trên như được minh họa trong Hình 2 với truy vấn "*hoa tình yêu*".



Hình 1: Tìm kiếm với từ khóa "*apple*"

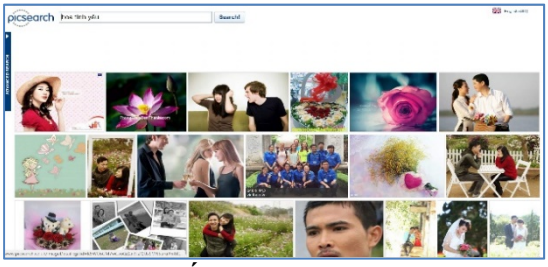


Hình 2: Tìm kiếm với từ khóa "*hoa tình yêu*"

Picsearch là công cụ chuyên tìm kiếm ảnh bằng cách chỉ mục các ảnh trên web với hơn 3 tỷ ảnh. Người dùng có thể nhập từ khóa cần tìm kiếm vào và hệ thống sẽ tìm kiếm ảnh chủ yếu dựa trên các meta-data đi kèm với ảnh. Cũng tương tự như hai bộ máy tìm kiếm trên, Picsearch cũng gặp những vấn đề như đã đề cập bên trên như được minh họa trong Hình 3.

Khác với các công cụ tìm kiếm trên, Incogna là một công cụ tìm kiếm ảnh dựa trên nội dung. Các ảnh trong bộ máy tìm kiếm này được phân lớp sẵn dựa trên nội dung của ảnh và người dùng có thể tìm

ảnh dựa vào nội dung. Do đó, công cụ tìm kiếm này có thể khắc phục các hạn chế của các bộ máy tìm kiếm trên. Tuy nhiên, công cụ này vẫn đang trong quá trình thử nghiệm. Hình 4 minh họa kết quả tìm kiếm trên Incogna với từ khóa "*obama family*".



Hình 3: Tìm kiếm với từ khóa "*hoa tình yêu*"

Hình 4: Kết quả tìm kiếm với từ khóa "*obama family*"

Trong nghiên cứu của Magesh và Thangaraj đã đề xuất một phương pháp tìm kiếm ảnh bằng nội dung dựa trên các mô tả được định nghĩa bằng ngôn ngữ RDF (Resource Description Framework) gán kèm theo mỗi ảnh (Liu *et al*., 2007; Magesh and Thangaraj, 2011). Các câu truy vấn ảnh của người dùng sẽ được biến đổi về cú pháp của SPARQL để truy vấn hình ảnh được mô tả bằng RDF.

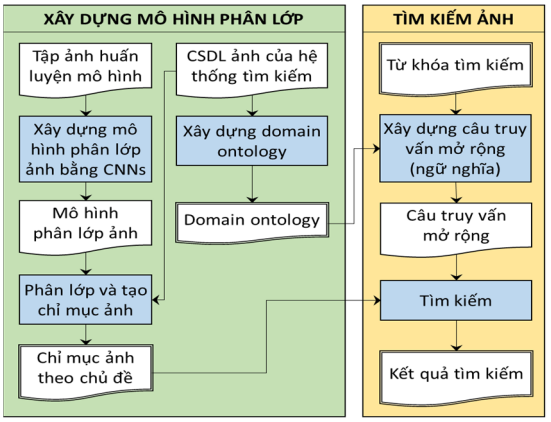
Trong nghiên cứu của Hyvönen *et al*. (2003), các tác giả đã trình bày một phương pháp tìm kiếm ảnh theo ngữ nghĩa bằng cách kết hợp meta-data đi kèm với ảnh và ontology của công nghệ web ngữ nghĩa. Ưu điểm của phương pháp này là dùng ontology để tạo một mạng ngữ nghĩa cho những thông tin có liên quan đến các ảnh trong bộ máy tìm kiếm. Do đó, phương pháp tìm kiếm này có thể gợi ý những hình ảnh có liên quan về ngữ nghĩa ngoài các kết quả tìm kiếm dựa trên meta-data.

Nghiên cứu gần đây của Shabaz Basheer Patel và Anand Sampat đã sử dụng kỹ thuật học sâu theo sự kết hợp giữa mạng CNNs để phân lớp ảnh và mạng RNNs để phân tích ngôn ngữ tự nhiên câu truy vấn nhằm xây dựng hệ thống tìm kiếm ảnh bằng ngôn ngữ tự nhiên (Patel and Sampat, 2017). Phương pháp này có ưu điểm là việc tìm kiếm không cần meta-data. Việc sử dụng CNNs còn giúp tận dụng được ưu điểm của công nghệ học sâu trong phân lớp nội dung ảnh. Ngoài ra, kết hợp ngôn ngữ tự nhiên trong tìm kiếm giúp cho người dùng có thể đưa ra các truy vấn tìm kiếm một cách tự nhiên, gần gũi hơn.

3 HỆ THỐNG TÌM KIẾM ẢNH THEO NỘI DUNG VÀ NGỮ NGHĨA

3.1 Kiến trúc của hệ thống

Trong nghiên cứu này, hệ thống tìm kiếm ảnh sẽ không hỗ trợ tìm kiếm theo dạng ngôn ngữ tự nhiên mà chỉ hỗ trợ người dùng tìm theo từ khóa hoặc nội dung ảnh truy vấn theo những chủ đề ảnh đã định trước. Tìm theo nội dung ở đây có nghĩa là nhãn của mỗi ảnh sẽ được gán dựa trên nội dung của ảnh thông qua mô hình phân lớp CNNs. Kiến trúc của hệ thống tìm kiếm ảnh theo nội dung kết hợp với ngữ nghĩa được trình bày trong Hình 5. Hệ thống này được xây dựng dựa trên mô hình phân loại ảnh CNNs và kết hợp với domain ontology để hỗ trợ tìm kiếm theo ngữ nghĩa.

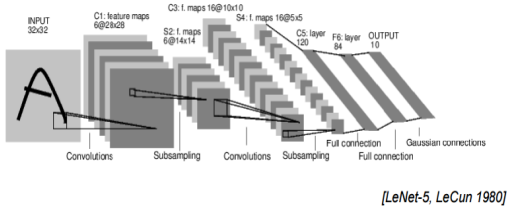


Hình 5: Kiến trúc của hệ thống

3.2 Xây dựng mô hình phân lớp ảnh bằng CNNs

Mạng neural tích chập (Convolutional Neural Network - CNNs) được gọi tắt là ConvNet (Krizhevsky *et al*., 2012) là một dạng của mạng nơron đa tầng, mỗi tầng thuộc một trong 3 dạng: tích chập (convolution), lấy mẫu con (subsampling), kết nối đầy đủ (full connection) được mô tả trong Hình 6.

CNN xem ảnh đầu vào là tầng input, mỗi pixel là một nơron, ảnh đầu vào này còn gọi là feature map. Feature map có thể coi là một ảnh thông thường, trong đó mỗi pixel được gọi là một nơron.

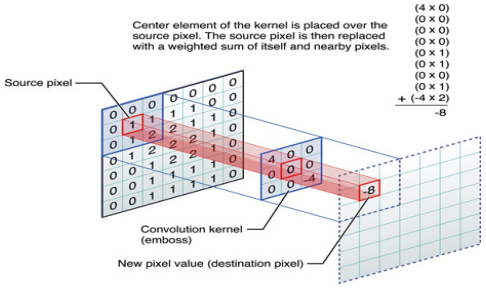


[LeNet-5, LeCun 1980]

Hình 6: Mô hình CNN nhận dạng chữ viết tay

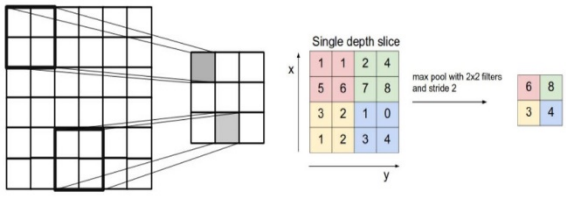
Tầng tích chập (C- convolution): hoạt động như bộ trích chọn đặc trưng, nghĩa là một hay nhiều kết xuất của tầng trước được tích chập với một hay

nhiều kernel để sinh ra một hay nhiều kết xuất (feature map) và được mô tả qua Hình 7.



Hình 7: Minh họa tích chập

Tầng lấy mẫu con (S - subsampling): Lấy mẫu con của mạng nơron tích chập giúp mạng chịu được những biến dạng của dữ liệu như tịnh tiến, quay, nghiêng. Toán tử lấy mẫu con được thể hiện trong Hình 8.



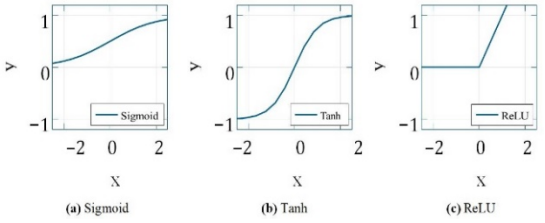
Hình 8: Minh họa việc lấy mẫu con

Tầng kết nối đầy đủ (F - Full connection): thực hiện công việc phân lớp như trong mạng nơron đa lớp thông thường,... Các tầng sau, mỗi tầng có một số feature map, mỗi feature map có một số filter (kernel) duy nhất, số lượng kernel bằng số lượng feature map ở tầng trước. Các kernel trong cùng một feature map có kích thước bằng nhau, kích thước kernel là một tham số của bài toán thiết kế mạng. Các giá trị điểm ảnh trong một feature map được tính toán bằng tổng các tích chập của các kernel tương ứng với các feature map trong tầng trước. Số lượng feature map trong tầng cuối cùng (tầng output) bằng số lượng kết xuất đầu ra của bài toán. Ví dụ: trong bài toán nhận dạng các số từ 0 đến 9, thì sẽ có 10 feature map trong tầng output và feature map nào có giá trị cao nhất sẽ được dùng làm kết quả của bài toán.

Trong mô hình Feedforward Noron Network (mạng nơron truyền thống), các layer kết nối trực tiếp với nhau thông qua trọng số w. Trong mô hình CNN thì các layer được kết nối với nhau thông qua cơ chế convolution. Nghĩa là nơron ở layer phía sau kết nối với nơron ở layer phía trước thông qua filter (chứ không kết nối trực tiếp với nơron phía trước).

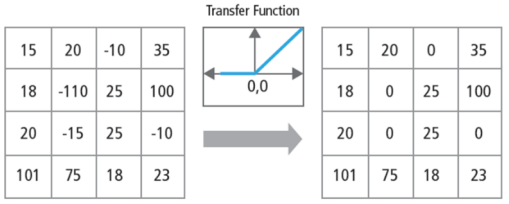
Thao tác trong mỗi nơron là tính giá trị đầu ra tương ứng với giá trị đầu vào f thông qua hàm kích hoạt hay còn gọi là hàm chuyển (hàm kích hoạt) g(f).

Một số hàm chuyển thường được sử dụng được liệt kê ở Hình 9.



Hình 9: Các hàm chuyển/kích hoạt

Hàm kích hoạt được sử dụng trong nghiên cứu này là hàm ReLU có chức năng được mô tả trong Hình 10.



Hình 10: Minh họa chức năng ReLU

Nghiên cứu này sử dụng hàm Momentum để tối ưu hóa độ lỗi cho mô hình (Qian, 1999). Phương pháp này sẽ điều chỉnh các vector trọng số theo cả hai bước lặp hiện tại và bước lặp trước đó. Phương pháp này được biểu diễn theo phương trình sau:

$$\Delta \mathbf{w}_t = -\epsilon \nabla_{\mathbf{w}} E(\mathbf{w}) + p\Delta \mathbf{w}_{t-1}$$

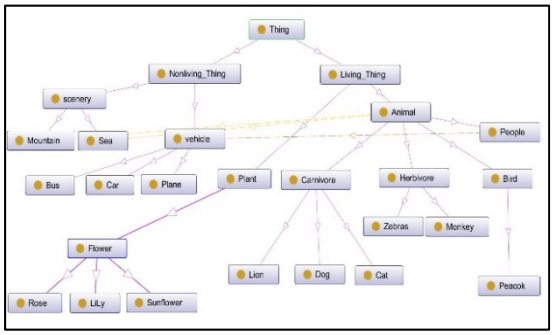
Trong đó, p là tham số momentum, $\nabla_{\mathbf{w}}$ là thao tác gradient ứng trọng số w, ϵ là learning rate.

3.3 Tìm kiếm theo ngữ nghĩa

Trong hệ thống này, chúng tôi dùng ontology làm cơ sở cho việc kết hợp ngữ nghĩa vào tìm kiếm. Ontology là một phương thức biểu diễn tri thức chuẩn cho web ngữ nghĩa (Liu *et al*., 2008). Phương thức biểu diễn tri thức này cho hình hóa các khái niệm và quan hệ giữa các khái niệm trong miền tri thức, cho phép các tri thức có thể được sử dụng lại cũng như được chia sẻ giữa các ứng dụng (Li *et al*., 2005; Lindén *et al*., 2004).

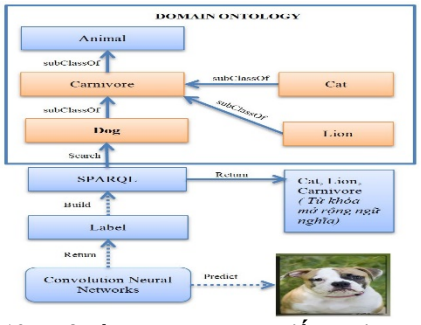
Dựa vào CSDL ảnh của hệ thống, chúng tôi xây dựng ontolgy cho một miền tri thức (domain-ontology) của các chủ đề ảnh để mô tả các khái niệm, các mối quan hệ ngữ nghĩa giữa chúng. Một ontology có thể được trực quan hóa bằng một đồ thị có hướng như Hình 11 với các đỉnh là các khái niệm và các cạnh biểu diễn mối quan hệ giữa các khái niệm. Nghiên cứu này đã xây dựng một domain-ontology bao gồm 15 khái niệm liên quan đến chủ đề của các ảnh trong CSDL. Onotlogy này có thể

được mở rộng tương ứng với tập dữ liệu ảnh trong CSDL.



Hình 11: Domain ontology của hệ thống

Các mối quan hệ được biểu diễn trong miền tri thức này bao gồm quan hệ về cha con như những khái niệm *Cat*, *Dog*, *Lion* đều là lớp con của lớp động vật ăn thịt (carnivore). Các khái niệm *Car*, *Bus*, *Plane* là lớp con của lớp phương tiện (vehicle) và các khái niệm *Sunflower*, *Rose*, *LiLy* là lớp con của lớp hoa (*Flower*). Ngoài ra, các khái niệm *Sea*, *Mountain* đều là con của lớp *Scenery* và một số mối quan hệ khác,...



Hình 12: Mở rộng từ khóa tìm kiếm với ngữ nghĩa

Trong hệ thống này, để thực hiện tìm kiếm theo ngữ nghĩa thì bước đầu tiên là thực hiện mở rộng từ khóa tìm kiếm. Bước này được hiện bằng cách xây dựng câu truy vấn SPARQL thích hợp từ các từ khóa tìm kiếm và thực hiện câu truy vấn trên domain-ontology. Sau đó, các từ khóa mở rộng được sử dụng để tìm kiếm các ảnh đã được chỉ mục trong hệ thống. Hình 12 minh họa thao tác mở rộng kết quả tìm kiếm bằng cách sử dụng CNNs để tìm ra nhãn của ảnh truy vấn của người dùng. Sau đó từ khóa mô tả nhãn của ảnh sẽ dùng để xây dựng câu truy vấn SPARQL. Cuối cùng sử dụng câu truy vấn SPARQL để tìm và trả về thêm các từ khóa có liên hệ về ngữ nghĩa trong domain-ontology.

4 THỰC NGHIỆM

4.1 Môi trường và các công cụ sử dụng cho thực nghiệm

Thực nghiệm được thực hiện trên 3 máy tính cài đặt theo mô hình Spark có cấu hình như sau:

Thành phần	Cấu hình
CPU	Intel(R) Core(TM) i7-2600 CPU @ 3.40GHz
RAM	8GB
OS	Linux
Bộ nhớ ngoài	120GB

Các thư viện và phần mềm hỗ trợ học sâu được sử dụng trong thực nghiệm là Miniconda, Tensoflow, TF-Learn và PyCharm.

4.2 Tập dữ liệu thực nghiệm

Tập dữ liệu thực nghiệm trong nghiên cứu này được thu thập từ trang web tìm kiếm ảnh Flick. Đây là một kho lưu trữ ảnh lớn, uy tín với hơn 10 tỷ ảnh có độ phân giải tốt. Có tất cả 40.803 ảnh được thu thập, bao gồm 15 chủ đề là: Cat, Dog, Peacok, LiLy, Car, Mountain, Sea, Sunflower, Plane, Rose, Lion, Zebras, Bus, Monkey và People. Các ảnh được điều chỉnh lại theo cùng độ phân giải là 64x64.

4.3 Xây dựng mô hình phân loại ảnh cho hệ thống tìm kiếm

Để xây dựng mô hình phân loại ảnh cho hệ thống tìm kiếm, tập dữ liệu thực nghiệm được chia thành 3 tập dữ liệu con là tập huấn luyện (training set) gồm 24.481 ảnh (60%), tập kiểm thử (test set) gồm 8.160 ảnh (20%) và tập giám sát (validation set) gồm 8.160 ảnh (20%). Tập dữ liệu giám sát dùng để giám sát quá trình học xem mạng có đang trong trạng thái bị học chậm (underfitting) hoặc quá khớp (overfitting) không?

Dữ liệu huấn luyện được chia thành từng batch, với *batch size* là 500 ảnh để đưa vào huấn luyện nhằm tránh tắt nghẽn mạng và giảm dung lượng bộ nhớ cần thiết để huấn luyện. Phương pháp tính độ lỗi của mạng là Momentum với các tham số *base learning* là 0,05 và *lr_decay* là 0,96. Khi mạng đã học qua toàn bộ ảnh trong tập huấn luyện một lần thì được xem như mạng đã học được một chu kỳ (epoch) và số epoch là 400 nên phải lặp tối đa là 19.584 lần theo công thức như sau:

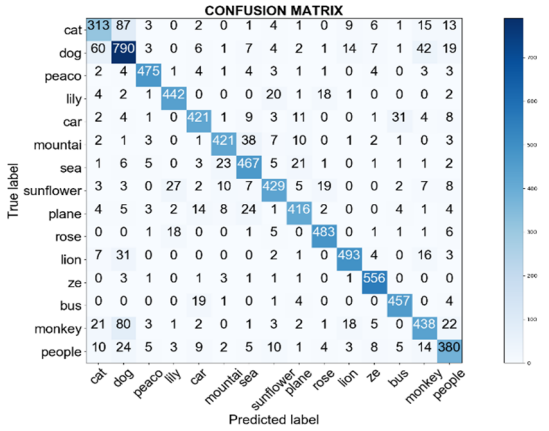
$$Số\ lần\ lặp = \frac{\text{Số mẫu train}}{\text{batch size}} \times số\ epoch$$

Qua quá trình huấn luyện và kiểm tra mô hình mạng neural tích chập đã thu được một số kết quả qua bảng thống kê về thời gian huấn luyện mô hình ở Bảng 1.

Để học được mô hình đạt độ chính xác là 83% trên tập train thì cần hơn 20 giờ huấn luyện. Để kiểm tra độ chính xác của mô hình phân lớp trên tập test thì chúng tôi đánh giá theo 3 độ đo là Precision ở Hình 14 và Confusion matrix ở Hình 13.

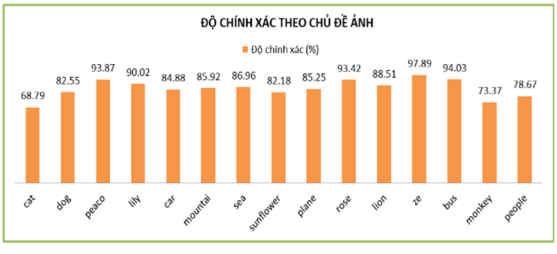
Bảng 1: Kết quả huấn luyện mô hình

Thời gian (giờ:phút:giây)	Bước lặp	Độ chính xác
00:13:26	180	31%
01:27:37	1.217	63%
02:54:21	2.397	69%
03:25:00	2.826	72%
04:35:15	3.774	74%
05:33:49	4.589	75%
06:26:10	5.313	77%
08:08:00	6.730	79%
11:37:34	9.683	80%
12:28:33	10.360	81%
19:28:00	16.180	82%
20:30:00	19.584	83%



Hình 13: Confusion matrix theo các chủ đề ảnh

Hình 13 và Hình 14 cho thấy mô hình phân lớp chủ đề ảnh ngựa vằn (zebras) có độ chính xác khá cao, đạt 97,89%. Nguyên nhân là do các đặc trưng của ngựa vằn nổi bật hơn so với các chủ đề ảnh còn lại như có các vệt vằn đen trên lưng. Ngược lại mô hình phân lớp chủ đề ảnh mèo (cat) có độ chính xác chưa cao, chỉ đạt 68,79%. Nguyên nhân có thể là do các đặc trưng ở mèo khó nhận dạng hơn các chủ đề ảnh khác và có nhiều đặc trưng tương đồng với một số chủ đề ảnh khác như chó và khỉ vì chúng đều là những động vật 4 chân, hình dáng đều nhỏ nhắn và chỉ khác biệt rõ nhất ở gương mặt.



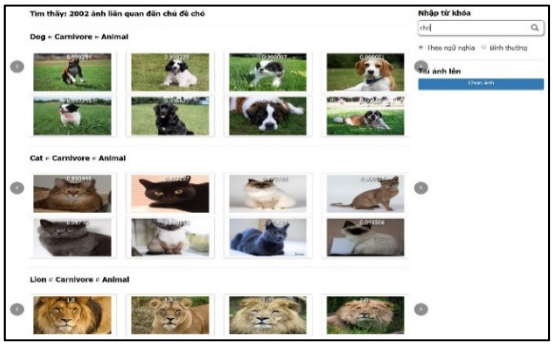
Hình 14: Biểu đồ độ đo Precision phân lớp theo các chủ đề ảnh

4.4 Kết quả tìm kiếm ảnh theo ngữ nghĩa

Ứng dụng tìm kiếm ảnh này được xây dựng trên nền tảng web, dựa vào kiến trúc đã mô tả trong Phần 3.3. Để minh họa cho chức năng tìm kiếm ngữ nghĩa, ứng dụng này cho phép lựa chọn tìm kiếm có hoặc không có ngữ nghĩa. Ví dụ, khi tìm kiếm với từ khóa "*dog*" (con chó), nếu tắt chức năng ngữ nghĩa thì kết quả chỉ trả về những hình ảnh được xếp loại là "*dog*", giống như các bộ máy tìm kiếm thông dụng khác.

Tuy nhiên, khi bật tính năng tìm kiếm kết hợp ngữ nghĩa cho thấy hệ thống không chỉ trả về những ảnh có chủ đề "*dog*" mà còn có thêm hai loài động vật nữa đó là mèo và sư tử. Kết quả thu được như trên là nhờ sự mở rộng thêm ngữ nghĩa cho từ khóa truy vấn bằng domain-ontology của hệ thống. Với sự suy luận trên ontology thông qua câu truy vấn SPARQL, mèo và sư tử cũng là loài động vật ăn thịt như loài chó nên các hình ảnh của 2 loài này sẽ được trả về trong kết quả truy vấn mở rộng. Kết quả được minh họa trong Hình 15 và Hình 16.

Hình 15: Tìm ảnh theo nội dung đã chọn về chó nhưng không theo ngữ nghĩa



Hình 16: Tìm theo ngữ nghĩa với từ khóa "*dog*"

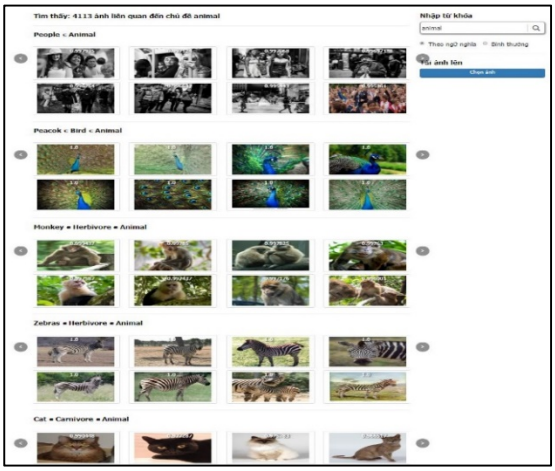
Hình 17 và Hình 18 minh họa trường hợp ngược lại với trường hợp trên, trong đó hệ thống sử dụng quan hệ hyponym để xác định các từ khóa mở rộng. Do không có hình ảnh nào trong hệ thống có chủ đề là "*animal*" nên khi tìm không ngữ nghĩa với từ khóa này sẽ không có ảnh nào tìm được. Tuy nhiên, khi sử dụng ngữ nghĩa thì sẽ trả về hình của một số loại động vật như chó, mèo, sư tử, khỉ, ngựa vằn, công,... vì tất cả các loài này đều là động vật.

Hình 17: Tìm kiếm không ngữ nghĩa với từ khóa "*animal*"

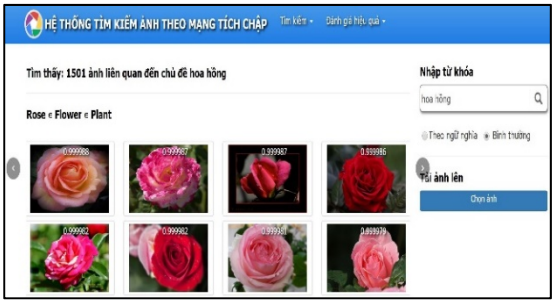
Ngoài ra, tính năng ngữ nghĩa hóa của hệ thống còn cho phép chuẩn hóa từ khóa truy vấn của người dùng nên khi tìm kiếm với từ khóa như "*hoa hồng*" hay "*hoa tình yêu*" thì hệ thống vẫn tìm kiếm được cùng một chủ đề ảnh như nhau được minh họa trong Hình 19 và Hình 20.

5 KẾT LUẬN

Nghiên cứu đã xây dựng thành công domain ontology, nó giúp cho việc biểu diễn mối quan hệ ngữ nghĩa giữa các chủ đề ảnh được rõ ràng hơn. Ngoài ra, nó còn giúp cho việc chuẩn hóa từ khóa tìm kiếm và đồng thời giúp mở rộng thêm kết quả tìm kiếm. Kết hợp với mô hình phân lớp ảnh có được qua quá trình huấn luyện bằng mạng neural tích chập đã giúp tìm thấy được những hình ảnh bị sai sót thông tin văn bản đi kèm (meta-data); từ đó cho thấy được tính khả thi của phương pháp xây dựng hệ thống tìm kiếm ảnh theo nội dung và ngữ nghĩa.



Hình 18: Tìm kiếm ngữ nghĩa với từ khóa "animal"



Hình 19: Kết quả tìm kiếm với từ khóa "*hoa hồng*"

Hình 20: Kết quả tìm kiếm với từ khóa "*hoa tình yêu*"

TÀI LIỆU THAM KHẢO

Hyvönen, Eero, Samppa Saarela, Avril Styrman, and Kim Viljanen. 2003. "Ontology-Based Image Retrieval." In WWW(Posters).

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. 2012. "Imagenet Classification with Deep Convolutional Neural Networks." In Advances in Neural Information Processing Systems, 1097–1105.

Li, Man, Xiao-Yong Du, and Shan Wang. 2005. "Learning Ontology from Relational Database." In Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference On, 6:3410–3415. IEEE.

Lindén, Krister, Jussi Olavi Piitulainen, and others. 2004. "Discovering Synonyms and Other Related Words." In Proceedings of COLING 2004 CompuTerm 2004: 3rd International Workshop on Computational Terminology.

Liu, Qiaoling, Kaifeng Xu, Lei Zhang, Haofen Wang, Yong Yu, and Yue Pan. 2008. "Catriple: Extracting Triples from Wikipedia Categories." In The Semantic Web, 330–44. Lecture Notes in Computer Science. Springer, Berlin, Heidelberg. doi:10.1007/978-3-540-89704-0_23.

Liu, Ying, Dengsheng Zhang, Guojun Lu, and Wei-Ying Ma. 2007. "A Survey of Content-Based Image Retrieval with High-Level Semantics." Pattern Recogn. 40 (1): 262–282.

Magesh, N., and P. Thangaraj. 2011. "Semantic Image Retrieval Based on Ontology and SPARQL Query." In International Conference on Advanced Computer Technology (ICACT).

Meeker, Mary. 2014. "Internet Trends 2014–code Conference." Retrieved May 28: 2014.

Patel, Shabaz Basheer, and Anand Sampat. 2017. "Semantic Image Search Using Queries." Accessed September 8.

Qian, Ning. 1999. "On the Momentum Term in Gradient Descent Learning Algorithms." Neural Networks 12 (1): 145–151.