Logo

Description automatically generated

BÁO CÁO ĐỒ ÁN

**ĐỀ TÀI**

**SEARCH PRODUCT**

**Môn học:** Thị giác máy tính trong tương tác người máy

(CS532.M21.KHCL)

**Giảng viên:** Đỗ Văn Tiến

Thành viên:

Nguyễn Dương Hải – 19521464

Đỗ Trọng Khánh – 19521676

Trịnh Tuấn Nam – 19521874

Nguyễn Thành Trung – 19522432



**Lời cảm ơn**

Đầu tiên, chúng em xin gửi lời cảm ơn và lòng biết ơn sâu sắc nhất tới TS. Đỗ Văn Tiến – giảng viên phụ trách bộ môn *“Thị giác máy tính trong tương tác Người-Máy”,* trực thuộc Khoa Khoa học Máy Tính, đã trang bị cho em những kiến thức trong môn học này.

Tuy nhiên trong quá trình nghiên cứu đề tài, do kiến thức chuyên ngành còn hạn chế nên chúng em vẫn còn nhiều thiếu sót khi tìm hiểu, đánh giá, trình bày về đề tài. Những góp ý của thầy trong buổi báo cáo là những kinh nghiệm quý báu mà chúng em học hỏi, khắc phục cho quá trình học tập và làm việc sau này.

Xin chân thành cảm ơn Thầy.

**Nội dung**

[**I TỔNG QUAN** 2](#_heading=h.gjdgxs)

[1. Ngữ cảnh ứng dụng bài toán 2](#_heading=h.30j0zll)

[2. Bài toán 3](#_heading=h.1fob9te)

[3. Ưu nhược điểm bài toán mang lại 3](#_heading=h.3znysh7)

[**II CÁC NGHIÊN CỨU TRƯỚC** 4](#_heading=h.2et92p0)

[1. Google Images Search 5](#_heading=h.tyjcwt)

[2. Bing 5](#_heading=h.3dy6vkm)

[**III NỘI DUNG CHÍNH** 6](#_heading=h.1t3h5sf)

[1. Xử lí hình ảnh 6](#_heading=h.4d34og8)

[2. Trích xuất đặc trưng của ảnh 6](#_heading=h.2s8eyo1)

[3. Cosine similarity 7](#_heading=h.17dp8vu)

[**IV THỰC NGHIỆM VÀ ĐÁNH GIÁ** 8](#_heading=h.3rdcrjn)

[**V KẾT LUẬN** 9](#_heading=h.26in1rg)

# **I TỔNG QUAN**

## Ngữ cảnh ứng dụng bài toán

Cùng với sự phát triển của các công nghệ kĩ thuật số và sự phổ biến của các thiết bị có chức năng quay phim, chụp ảnh. Nhu cầu sử dụng các chức năng quay phim chụp ảnh của người dùng cũng tăng lên. Những thiết bị có chức năng như rất phổ biến vì thế việc tìm kiếm thông tin dựa trên ảnh cũng đang được nghiên cứu và phát triển. Thay vì phải viết những câu truy vấn bằng văn bản, hay giọng nói thì việc chụp ảnh để tìm kiếm thông tin cũng đem lại nhiều sự tiện lợi cho người dùng.

Hầu hết người dùng đều sử dụng công cụ tìm kiếm bằng cách viết hoặc nói các câu truy vấn bằng văn bản để tìm kiếm thông tin của một sản phẩm. Việc sử dụng phương pháp này khá hiệu quả tuy nhiên vẫn tồn tại vài nhược điểm như người dùng có thể nhập sai nhập thiếu câu truy vấn, người dùng không biết các từ khóa liên quan đến sản phẩm, …. Từ những khó khăn trên phương pháp tìm kiếm bằng hình ảnh có thể khắc phục được phần nào các hạn chế của phương pháp tìm kiếm bằng truy vấn.

Phương pháp tìm kiếm bằng hình ảnh sẽ nhận đầu vào là hình ảnh sau đó sẽ trả ra tên của sản phẩm có trong ảnh.

## 2. Bài toán

**Text

Description automatically generated with medium confidence**

*Hình 1 Input và output bài toán*

## 3. Ưu nhược điểm bài toán mang lại

Phương pháp tìm kiếm bằng hình ảnh của google (Google Image Search) và một vài mạng xã hội như Flickr hay Pinterest hoạt động rất tốt. Bài toán gặp nhiều khó khăn khi thời gian tìm kiếm bằng hình ảnh là lâu hơn so với tìm kiếm bằng từ khóa. Tuy nhiên vẫn có một vài ưu điểm và phù hợp với một số lượng lớn người dùng nên bài toán tìm kiếm bằng hình ảnh vẫn là một bài toán có tiềm năng phát triển. Mục tiêu hướng đến là xây dựng một chương trình có thể trả về tên của sản phẩm chính xác và cố gắng để tốc độ nhận diện ra sản phẩm là tốt nhất có thể. Bài toán hướng đến các sản phẩm được bày bán trong siêu thị và sẽ chỉ giới hạn trong 40 sản phẩm thuộc các loại như (đồ ăn, đồ uống, gia vị, đồ gia dụng , …).

Graphical user interface, text, application

Description automatically generated

*Hình 2 Chức năng tìm kiếm bằng hình ảnh của google*

Đối với bài toán này sử dụng các phương pháp xử lí hình ảnh để hình ảnh thể hiện được sản phẩm một cách tốt nhất từ đó trích xuất các đặc trưng của hình ảnh và so sánh với các hình ảnh khác để cho ra được kết quả cuối cùng.

# **II CÁC NGHIÊN CỨU TRƯỚC**

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.

## 1. Google Images Search

***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.

Graphical user interface, website

Description automatically generated

*Hình 3 Kết quả tìm kiếm bằng hình ảnh và văn bản đi kèm*

## 2. Bing

***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.

# **III NỘI DUNG CHÍNH**

## Xử lí hình ảnh

Hình ảnh đầu vào sẽ được xử lí trước khi được lấy đặc trưng. Cụ thể:

* *Hình ảnh của sản phẩm sẽ được loại bỏ phần nền. Ở bước thực hiện này nhóm sử dụng thư viện REMBG(một công cụ xóa nền. Thư viện này sử dụng U2-net để nhận diện vật thể chính trong ảnh).*
* *Sau đó sẽ được cắt bỏ bớt phần nền xung quanh để hình ảnh cận sản phẩm nhất có thể(tìm kiếm tất cả các contour của hình ảnh, sau đó cắt bỏ theo contour có diện tích lớn nhất).*

A picture containing text, beverage, drink, can

Description automatically generated

*Hình 4: Quy trình xử lí hình ảnh*

## Trích xuất đặc trưng của ảnh

Hình ảnh sau khi được xử lí sẽ được trích xuất đặc trưng bằng một mạng CNN là VGG-16

*VGG-16 là một model phổ biến cho việc trích xuất đặc trưng của hình ảnh. VGG16 là mạng convolutional neural network được đề xuất bởi K. Simonyan and A. Zisserman, University of Oxford. Model sau khi train bởi mạng VGG16 đạt độ chính xác 92.7% top-5 test trong dữ liệu ImageNet gồm 14 triệu hình ảnh thuộc 1000 lớp khác nhau. VGG-16 bao gồm 13 layers tích chập 2 chiều và 3 layers fully connected.*

Waterfall chart

Description automatically generated

*Hình 5 Kiến trúc mạng VGG-16*

Sau các bước xử lí ảnh sẽ được đi qua VGG-16 để thu được một vector. Vector này thể hiện được đặc trưng của bức ảnh.

## Cosine similarity

Khi đã có được vector thể hiện đặc trưng của ảnh tiến hành so sánh vector của ảnh này với ảnh khác. Ở trong bài toán đã sử dụng cosine để tính góc giữa hai vector từ đó tìm ra hai vector gần giống nhau nhất(hai bức ảnh càng giống nhau, giá trị cosine sẽ càng lớn).

Độ đo cosin là một cách đo độ tương tự giữa 2 vector khác 0. Độ đo này được định nghĩa bằng giá trị cosine của góc giữa 2 vector. Độ đo này là một thẩm định có tính định hướng chứ không phải về độ lớn. Giá trị của độ đo này sẽ nằm trong khoảng từ [-1,1]. Độ đo có giá trị -1 khi 2 vector tạo thành góc 180o và 1 khi 2 vector tạo thành góc 0o. Điều này chứng minh giá trị cosine càng lớn thì góc giữa 2 vector càng nhỏ.

Text

Description automatically generated

*Hình 6 Công thức tính cosine similarity*

Hình 6 thể hiện công thức tính độ đo cosine. Với A và B là 2 vector (Ai và Bi lần lượt là các thành phần của 2 vector A, B).

# **IV THỰC NGHIỆM VÀ ĐÁNH GIÁ**

Tập dữ liệu: Tiến hành đi thu thập dữ liệu bằng cách chụp ảnh lại các sản phẩm trong siêu thị. Số sản phẩm được giới hạn lại 41 sản phẩm(các sản phẩm thuộc các loại sản phẩm như thức ăn, đồ uống, đồ gia dụng, đồ cá nhân, ….). Mỗi sản phẩm chụp khoảng 27 tấm ảnh. Tổng số lượng ảnh là 1098 ảnh.

Độ đo: Accuracy/ mAP (vì đây là bài toán truy vấn ảnh – Image Retrieval).

Bộ dữ liệu được chia làm 89%, 11%. Phần lớn hơn sẽ được xử lí, trích xuất đặc trưng sau đó được lưu trữ lại. 11% còn lại đối với mỗi ảnh cũng sẽ được xử lí, trích xuất đặc trưng sau đó so sánh với các ảnh đã được trích xuất đặc trưng trước đó và tìm ra vector đã được lưu trữ tạo với nó cosine similarity lớn nhất. Từ đó có thể suy ra được vector gần giống với ảnh nhất và trả về nhãn của ảnh.

Kết quả: Accuracy (86.17%) - mAP

Tuy cho hiệu suất khá tốt nhưng bù lại thời gian thực thi chương trình vẫn còn hạn chế, dưới đây là một số trường hợp dự đoán còn sai:



*Hình 7 Một số trường hợp dự đoán sai.*

Có thể thấy trong các trường hợp dự đoán sai trên thì đối tượng được dự đoán và đối tượng dự đoán có hình dạng (dạng lon/chai) và kích thước khá tương đồng dẫn đến kết quả dự đoán không đúng.

# **V KẾT LUẬN**

Đây là một bài toán truy vấn thông tin từ ảnh.

Sử dụng mô hình pre-train của *VGG16* để lấy các weights phục vụ cho dự đoán thông tin sản phẩm bằng phương pháp *Cosine Similarity* đã mang lại kết quả khá tốt.

Tuy nhiên, việc chọn mô hình để rút trích đặc trưng của nhóm còn mang tính chủ quan, chưa có phép so sánh với với các mô hình khác nên dù VGG16 mang lại kết quả khá tốt nhưng chưa nhận định được nó là một mô hình tối ưu cho bài toán này.

Tài liệu tham khảo