

CẢI THIỆN KHẢ NĂNG BIỂU DIỄN NGŨ NGHĨA CỦA HÌNH ẢNH TRONG MÔ HÌNH PIC2WORD BẰNG PHƯƠNG PHÁP SINH CHÚ THÍCH HÌNH ẢNH

Nguyễn Trần Việt Anh - MSSV: 21520006

¹ Trường ĐH Công nghệ thông tin, ĐH
Quốc gia TP Hồ Chí Minh

What ?

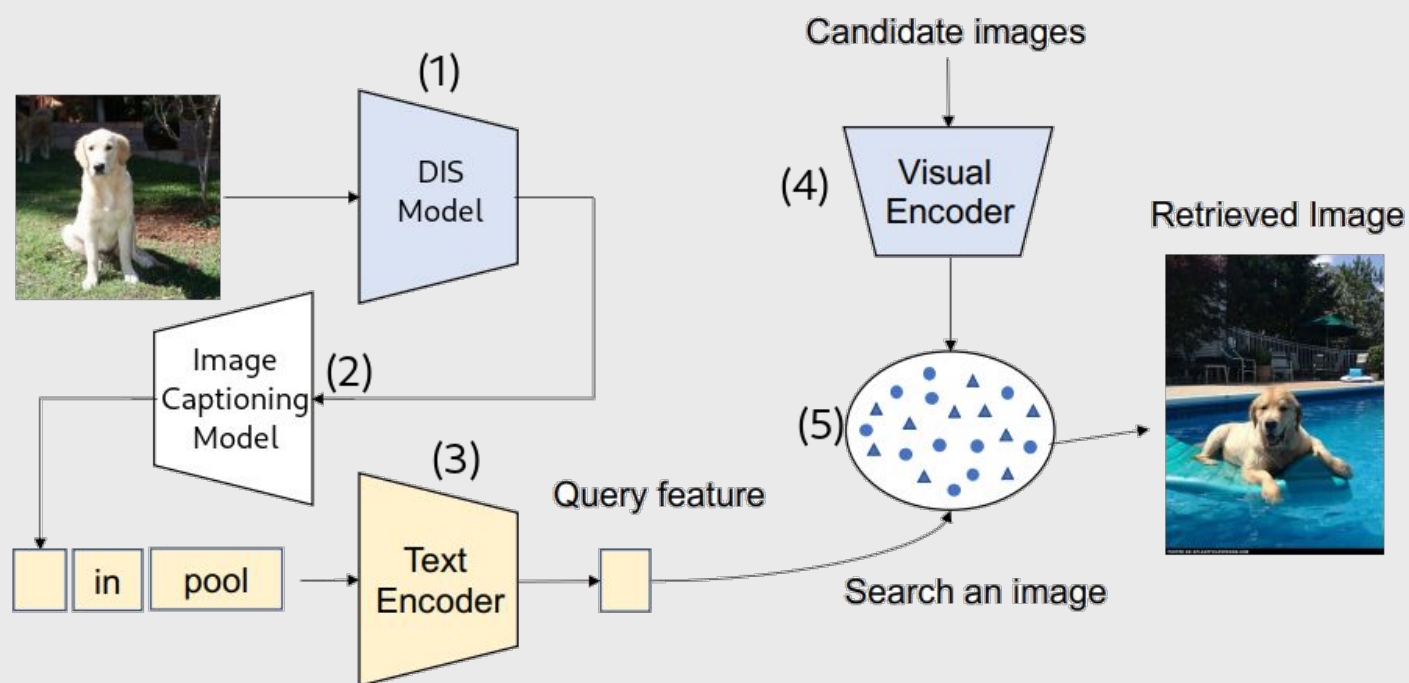
Chúng tôi đề xuất một phương pháp truy xuất hình ảnh hỗn hợp không cần nhãn trước bằng phương pháp sinh chú thích hình ảnh, cụ thể:

- Xây dựng một mô hình để truy vấn hình ảnh hỗn hợp không cần nhãn trước qua việc cải tiến mô hình Pic2Word.
- Huấn luyện, đánh giá mô hình trên các tập dữ liệu và độ đo khác nhau.

Why ?

- Trong cuộc sống hiện nay, sự phổ biến của ảnh số từ nhiều nguồn khác nhau như ảnh chụp từ điện thoại di động, hình ảnh từ mạng xã hội, website,... Đã dẫn tới nhu cầu rất lớn trong việc truy xuất, tìm kiếm hình ảnh.
- Bài toán truy vấn hình ảnh hỗn hợp không cần nhãn trước (ZS-CIR) là một bài toán mới, còn nhiều thách thức trong việc biểu diễn các thông tin về mặt ngữ nghĩa của hình ảnh và kết hợp các thông tin từ mô tả đầu vào và hình ảnh đầu vào.

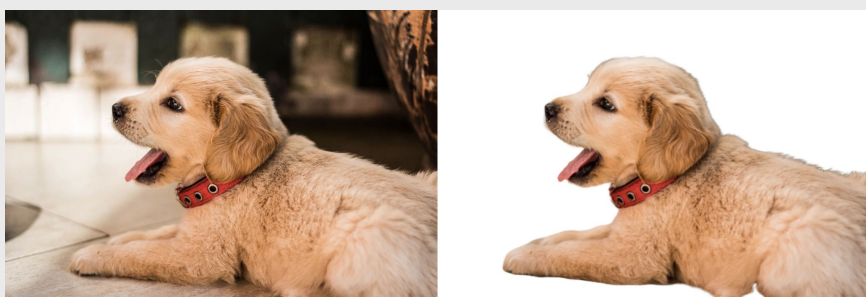
Overview



Description

1. Loại bỏ background của ảnh

- Chúng tôi sử dụng một mô hình pretrained InSPyReNet hoặc IS-Net để xóa background của ảnh, loại bỏ các thông tin không cần thiết.



2. Trích xuất mô tả của ảnh

- Một mô hình pretrained Image Captioning như mPLUG, OFA hoặc GIT được sử dụng để trích xuất ra đoạn mô tả của ảnh từ ảnh đầu vào.
- Đoạn mô tả này sau đó sẽ được kết hợp với mô tả đầu vào, tạo thành một đoạn mô tả kết hợp thông tin của hình ảnh và mô tả đầu vào.

3. Trích xuất Text Embedding Vector

- Chúng tôi sử dụng Text Encoder của mô hình pretrained CLIP để ánh xạ đoạn mô tả sang một embedding vector.

4. Trích xuất các đặc trưng từ các ảnh ứng viên

- Tất cả các hình ảnh ứng viên được đưa vào Visual Encoder của mô hình pretrained CLIP để ánh xạ hình ảnh sang các embedding vector.

5. So sánh các hình ảnh ứng cử viên

- So sánh embedding vector được trích xuất từ Text Encoder với mỗi embedding vector được trích xuất từ Visual Encoder để tìm ra những hình ảnh tương đồng nhất với truy vấn.