

NÂNG CAO KHẢ NĂNG TRUY XUẤT HÌNH ẢNH DỰA TRÊN CNN: KHẢO SÁT CÁC KỸ THUẬT ĐỂ CẢI THIẾN HIỆU SUẤT

Đào Tuấn Anh^{1,2}

Cao Thanh Bình^{1,3}

Đặng Phi Hùng^{1,4}

Nguyễn Hữu Tân^{1,5}

¹ Trường Đại học Công nghệ Thông tin
ĐHQG TP.HCM

{² 19520377, ³ 19520408, ⁴ 19520573, ⁵ 19520921}@gm.uit.edu.vn

What ?

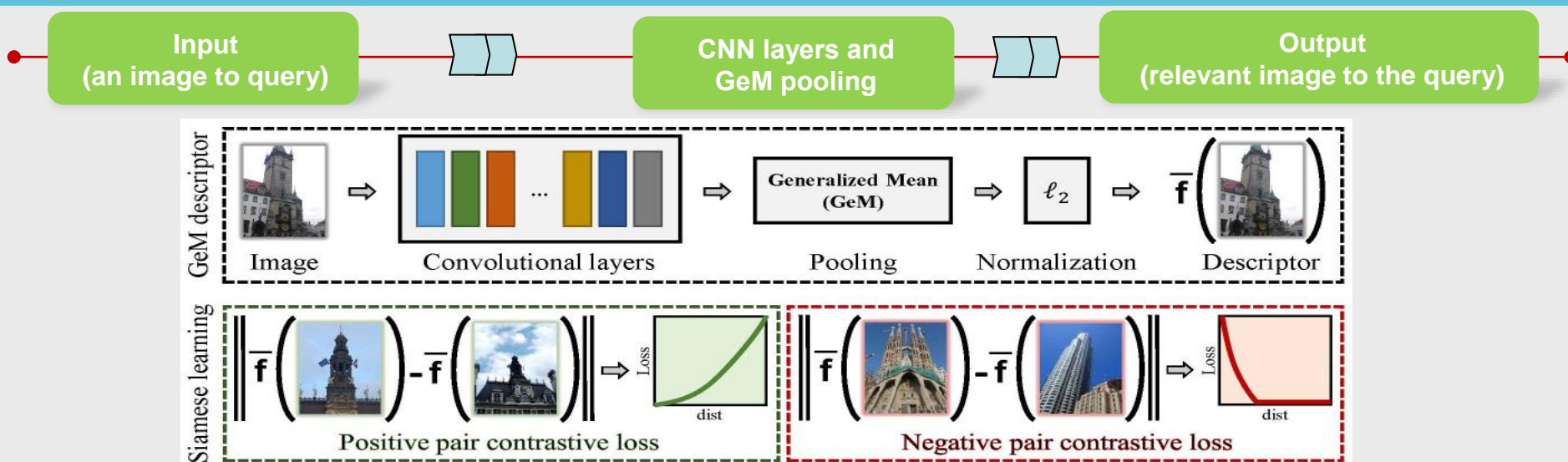
Chúng tôi dự kiến giới thiệu một phương pháp nâng cao khả năng truy xuất hình ảnh dựa trên CNN, trong đó bao gồm:

- Giới thiệu một lớp tổng hợp mới, biểu diễn hình ảnh đa quy mô và phương pháp mở rộng truy vấn cho việc tìm kiếm hình ảnh.
- Mở rộng nghiên cứu trước đó bằng cách tiến hành thêm thử nghiệm.

Why ?

- CNNs đã trở thành giải pháp cho vấn đề truy xuất hình ảnh sau khi Krizhevsky sử dụng ImageNet. Một cách tiếp cận khác để tiết kiệm thời gian hơn là tinh chỉnh mạng phân loại được train từ trước.
- Chúng tôi dự định áp dụng phương pháp tinh chỉnh không giám sát để truy xuất hình ảnh bằng CNN kết hợp các thông tin từ kỹ thuật SfM để train các CNN của riêng mình, sử dụng một pooling layer có thể train để khái quát hoá các sơ đồ gộp phổ biến hiện có cho CNN.

Tổng quan



Mô tả

1. Mạng nơ-ron tích chập (CNNs)

- CNN có khả năng ghi lại sự phụ thuộc không gian của hình ảnh kể từ khi nó xử lý chúng dưới dạng ma trận và phân tích toàn bộ các phần của một hình ảnh tại một thời điểm, tùy thuộc vào kích thước của bộ lọc.
- Các lớp liên kết được với nhau thông qua cơ chế tích chập. Lớp tiếp theo là kết quả phép tính tích chập từ lớp trước đó, nhờ vậy mà ta có được các kết nối cục bộ.
- Mỗi một lớp được sử dụng các filter khác nhau thông thường có hàng trăm hàng nghìn filter như vậy và kết hợp kết quả của chúng lại.

4. Hàm thất thoát

- Hàm mất mát contrastive loss là một phương pháp huấn luyện mạng để học cách so sánh và phân loại các đối tượng dựa trên sự tương đồng của chúng.
- Phát hiện của chúng tôi cho thấy rằng chức năng thất thoát tương phản khái quát hoá tốt hơn và hội tụ đến mức hiệu suất cao hơn.

2. GeM và Image Descriptor

- Generalized - Mean (GeM) là một phương pháp để tổng hợp các đặc trưng trên vector của một hình ảnh thành một vector đại diện duy nhất.
- Phương pháp tính toán trung bình có lũy thừa của các vector đặc trưng với các giá trị lũy thừa khác nhau, trong đó các giá trị lũy thừa được coi là các tham số có thể học được.
- Vector đầu ra thu được từ quá trình tổng hợp được chuẩn hoá bằng giai đoạn đánh giá, sản phẩm bên trong giữa hai hình ảnh được tính toán.

5. Làm trắng và kích thước

- Thành phần làm trắng thu được bằng cách lấy căn bậc hai nghịch đảo của ma trận phương sai C_S cho các cặp khớp với nhau.
- Phần quay được áp dụng bằng phương pháp Phân tích thành phần chính PCA vào ma trận phương sai của các cặp chưa từng có trong vùng làm trắng.

3. Siamese Learning

- Siamese learning là một loại học sâu liên quan đến việc đào tạo một mạng neural để so sánh hai đầu vào và xác định sự tương đồng của chúng.
- Một mạng lưới hai nhánh được train cho các công việc, dựa trên kiến trúc siamese. Cả hai nhánh của mạng lưới chia sẻ cùng một bộ tham số.
- Trong quá trình train, mạng lấy các cặp hình ảnh (i, j) làm đầu vào cùng với các nhãn tương ứng $Y(i, j) \in \{0, 1\}$ biểu thị xem cặp đó có khớp (0) hay chưa khớp.

6. Tái tạo mô hình 3D

- Kết hợp truy xuất hình ảnh Bag-of-Words (BoW) với Structure-from-Motion (SfM) bằng cách sử dụng quy trình truy xuất SfM hiện đại.
- Cách tiếp cận này hoàn toàn tự động và không yêu cầu đến bất kỳ sự can thiệp thủ công nào.