ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

KHOA: KHOA HỌC MÁY TÍNH

‒‒‒‒‒‒‒‒‒‒‒

BÁO CÁO ĐỒ ÁN MÔN HỌC

TRUY VẤN THÔNG TIN ĐA PHƯƠNG TIỆN

**IMAGE RETRIEVAL**



GVHD: Ngô Đức Thành

Sinh viên:

Phạm Hoàng Oanh – 16521762

Nguyễn Minh Châu – 16521553

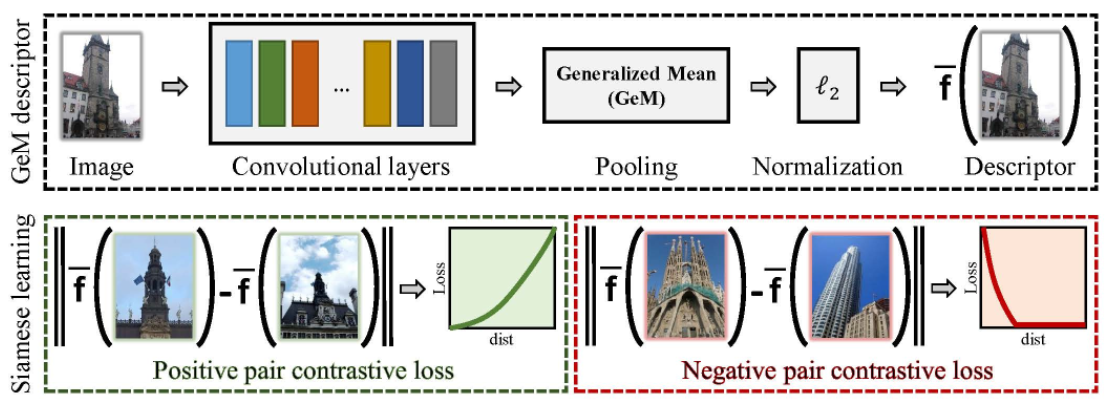
**Chương 1**

**Phương pháp thực hiện và các kỹ thuật được áp dụng**

* 1. **Phương pháp:**

Sử dụng mạng CNN đã được tinh chỉnh (fine-tunning) có làm trắng (whitening) để rút trích đặc trưng. Kết hợp với lớp Generalized-Mean (GeM) pooling [1][2].

Mạng CNN này ban đầu được dùng để phân loại ảnh sau đó được tinh chỉnh đáng kể để phù hợp với việc truy xuất ảnh. Đầu ra của mạng CNN này sẽ đi qua một lớp pooling mới (GeM) – có các tham số có thể học được, hoặc một tham số toàn cục hoặc trên một chiều của đầu ra.



Hình 1.1: Kiến trúc mạng được đề cập trong bài báo [1]. Một vector ¯f duy nhất được trích xuất để thể hiện một hình ảnh

Hàm loss:

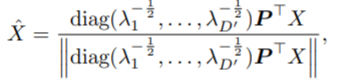
- Contrastive Loss (được cho là hội tụ tốt hơn Triplet Loss)

- Triplet Loss

* 1. **Các kỹ thuật áp dụng:**

*Làm trắng (Whitening):*

Là một phép biến đổi tuyến tính biến đổi một vector với các biến ngẫu nhiên với ma trận hiệp phương sai đã biết thành một tập hợp các biến mới không tương quan và mỗi biến có một phương sai. Với X là vector cho mô tả hình ảnh đầu vào (BOW hoặc VLAD) dự kiến cắt ngắn, làm trắng và tái chuẩn hóa thành một vector mới X’.



Hình 1.2: Công thức whitening được đề cập trong bài báo [3].

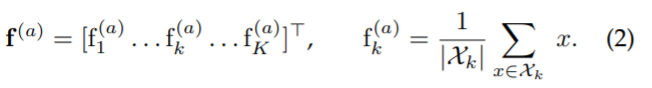
Trong đó, ma trận DxD’ P được hình thành bởi hàm riêng lớn nhất của ma trận hiệp phương sai C, và λi là giá trị riêng liên kết với vector riêng lớn nhất thứ i.

*Generalized-mean pooling:*

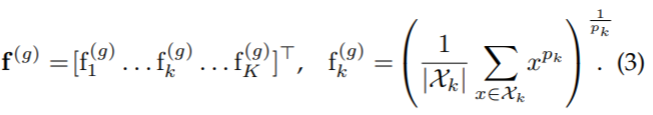
Thay vì là dùng global max pooling, lấy X làm đầu vào và tạo ra một vector f như sau:



Hay average pooling:



Thì sử dụng lớp tổng quát trung bình (Generalized-mean pooling)

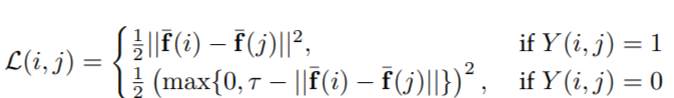


Pooling ở (1) và (2) là trường hợp đặc biệt của GeM pooling (3). Ví dụ max pooling khi pk  → ∞ và average pooling pk =1. Vector đặc trưng cuối cùng bao gồm một giá trị trên mỗi feature map, ví dụ geụ generalized-mean activation, và kích thước của vector có số chiều bằng K, đối với các mạng phổ biến thì K tương đương với 256, 512 hoặc 2048.

Tham số pk  có thể được thiết lập trước hoặc thông qua quá trình học.

Siamese learning và hàm Loss:

Áp dụng kiến trúc Siamese và huấn luyện trên mạng 2 nhánh. Mỗi nhánh là một bản sao của nhau, có nghĩa là dùng chung một tham số. Đầu vào gồm các cặp hình ảnh (i,j) và nhãn Y(i,j) ∈ {0, 1} khai báo một cặp là không khớp (nhãn 0) hay khớp (nhãn 1). Sử dụng hàm loss contrastive để so các cặp ảnh với nhau:



Trong đó, ¯f(i) là vector l2-normalized GeM của ảnh I,  là tham số lề xác định khi các cặp không khớp có khoảng cách đủ lớn để bỏ qua.

**Chương 2**

**Kết quả đánh giá hệ thống và phân tích**

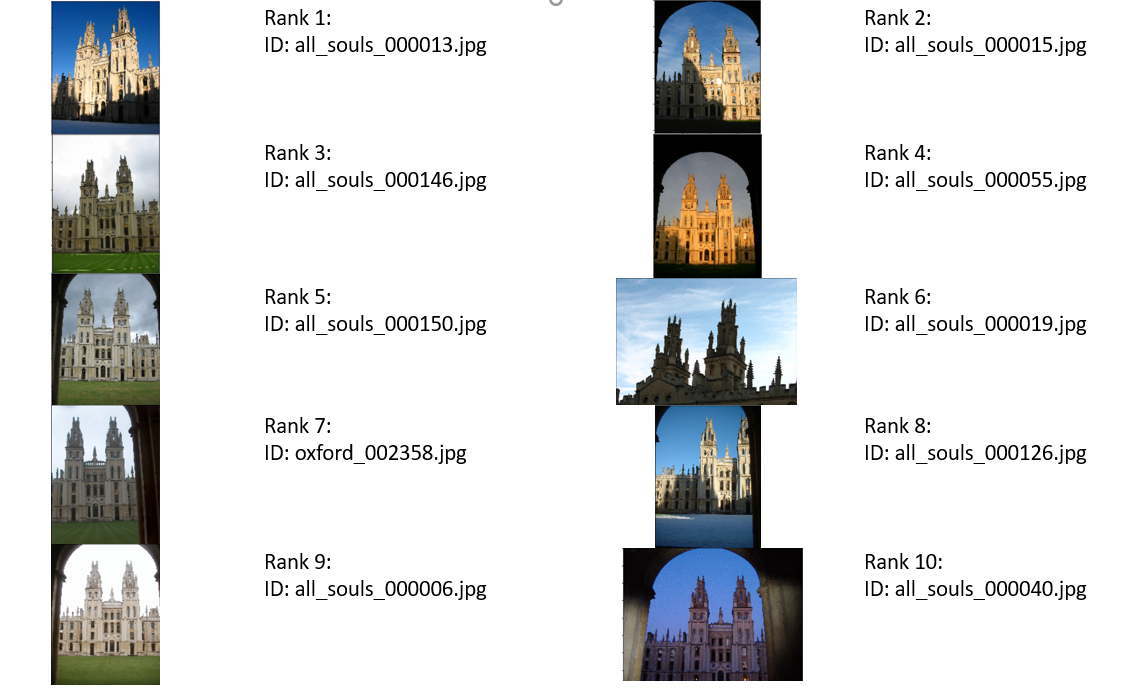
**2.1 Kết quả đánh giá hệ thống:**

Công thức tính Average Precision:

**2.2 Thử nghiệm:** Truy vấn một ảnh như sau:



Kết quả trả về là một danh sách các ảnh được cho là phù hợp với ảnh query:



Kết quả đánh giá trên độ đo mAP với query mẫu của dataset:

Oxford5k mAP: 79.23

Oxford5k + whiten mAP: 85.23

2.3 **Thời gian**

Để truy vấn 1 query với dataset Oxford5k thời gian chạy chưa đến 1s, nên hoàn toàn có thể chạy tương đối với các dataset có kích thước >100k. (Điều này cũng đã được khẳng định bởi tác giả).

**Tài liệu tham khảo:**

[1] Fine-tuning CNN Image Retrieval with No Human Annotation,Radenović F., Tolias G., Chum O., TPAMI 2018

[2] <https://github.com/filipradenovic/cnnimageretrieval-pytorch>

[3] Hervé Jégou, Ondrej Chum. Negative evidences and co-occurrences in image retrieval: the benefit of PCA and whitening. ECCV - European Conference on Computer Vision, Oct 2012, Firenze, Italy