VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY
FACULTY OF COMPUTER SCIENCE

FINAL PROJECT

IMAGE RETRIEVAL

Instructor

Ph.D. Thanh Duc Ngo

Students

Nhat Minh Phan - 19521956
Phat Vo Tien Le - 19521993
Trung Thanh Nguyen - 19522432
Vu Quang Hoang - 19522530

Class

CS336.N11.KHCL

Ho Chi Minh City - 2023

Contents

1 Introduction

Nowaday, with a huge amount of data, the demand of searching is also increase, not only search by text search but also images and sounds. This kind of searching can be called "Information Retrieval", it is the process of accessing and retrieving relevant information from a collection of data. Information retrieval is a crucial aspect of many activities and has become an integral part of our daily lives. In this report, we will introduce about a Image Retrieval.

Image Retrieval (IR) is a way of retrieving images from a database. In IR, a user specifies a query image and gets the images in the database relevant to the query image. To find the most relevant images, IR compares the content of the input image to the database images. More specifically, IR compares visual features such as shapes, colors, texture and spatial information and measures the similarity between the query image with the images in the database with respect to those features.

In this project, we build a Image Retrieval system and deploy it into a web app. The user chooses a image query into the system, the system will compute a the similarity of each image in the database matches the query, and rank the objects according to this value. The top ranking objects are then shown to the user. In this project we used two benchmark dataset, Oxford5k (5062 images) and Par6k (6412 images).

2 Dataset

The Oxford5k dataset consists of 5062 images collected from Flickr by searching for particular Oxford landmarks. We use this additional data to expand our collection to make it more challenge for the IR system.

The Par6k dataset, which is a collection of annotated images of buildings in Paris, France. This dataset was created by researchers at the University of California, Berkeley and was released in 2009. The Par6k dataset contains a total of 6,412 images of buildings in Paris, captured from various viewpoints and under different lighting conditions. The following 11 classes were used to collect the images from Flickr: La Defense Paris, Eiffel Tower Paris, Hotel des Invalides Paris, Louvre Paris, Moulin Rouge Paris, Musee d'Orsay Paris, Notre Dame Paris, Pantheon Paris, Pompidou Paris, Sacre Coeur Paris, Arc de Triomphe Paris.

Figure 1: Illustration of Oxford5k dataset

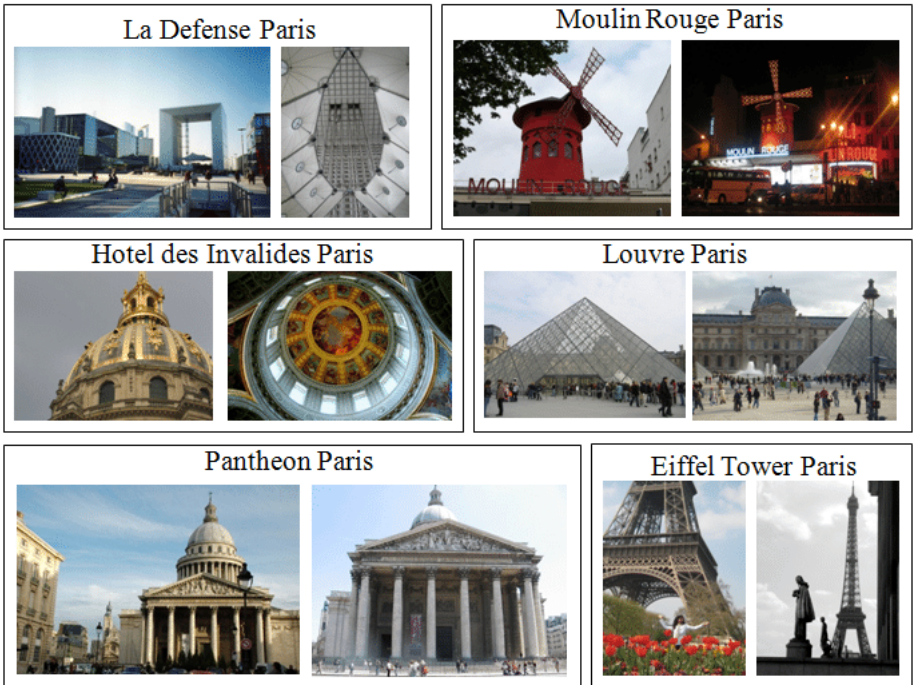

Figure 2: Illustration of Par6k dataset

We will evaluate this IR system using 55 images query in Par6k data set, we choose 5 image from each class. We choose query image base on the different building's viewpoint, lighting condition, distance from camera to the building and building is partially obscured by other objects.

3 Build IR system

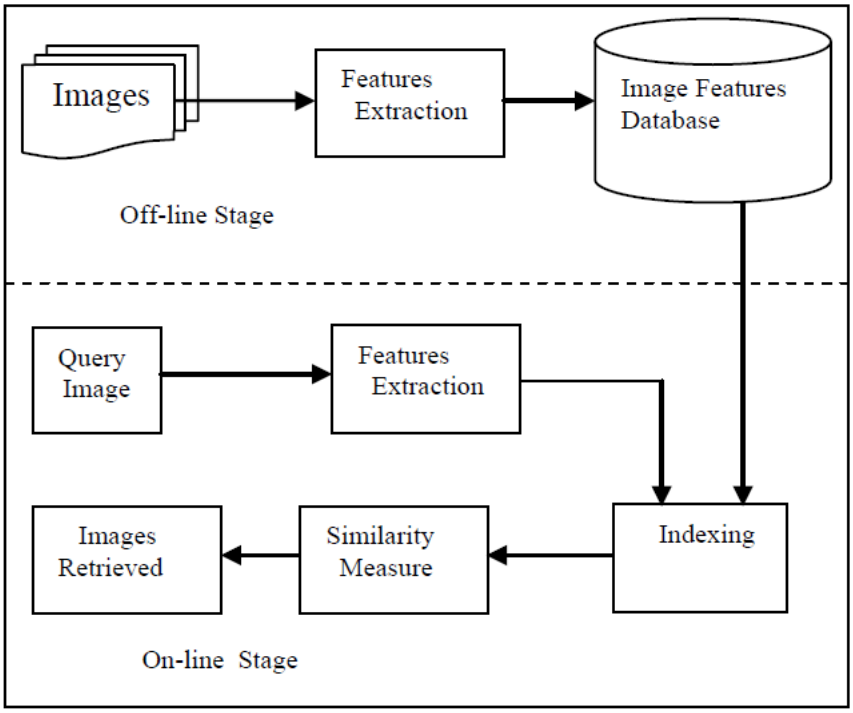

Figure 3: Illustration of our IR system

We try to build 3 feature extractor with pre-trained networks: VGG16, VGG19 and Resnet50 because we want to see the different in performance of three system. We choose this three networks because all of them are good networks with high performance on image problems.

VGG16 [**3**] is a deep convolutional neural network architecture that was proposed by the Visual Geometry Group at the University of Oxford . It was introduced in the 2014 ImageNet competition and achieved state-of-the-art results on the ImageNet dataset.

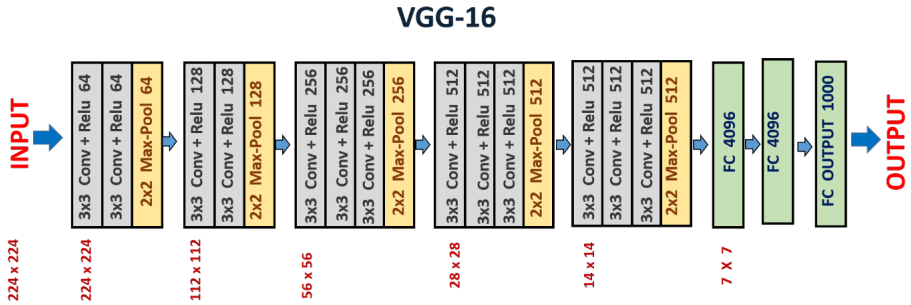

Figure 4: Illustration of VGG16 model

VGG19 [**4**] was an extension of VGG16 that was proposed the following year.The primary difference between VGG19 and VGG16 is the number of layers in each architecture. VGG19 has 19 layers, while VGG16 has 16 layers. This means that VGG19 has a deeper architecture than VGG16, which may lead to improved performance on some tasks, but can also increase the risk of overfitting or training instability.

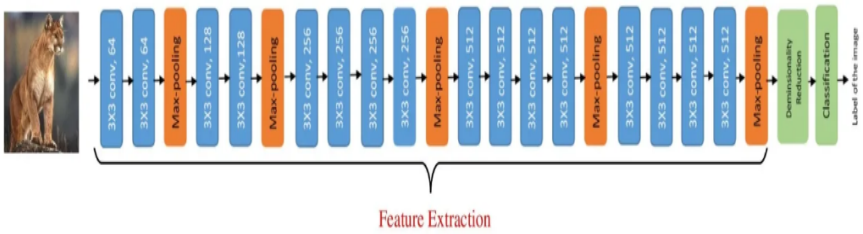

Figure 5: Illustration of VGG19 model

ResNet50 [**5**] is a deep neural network architecture that was introduced by Microsoft Research in 2015.The ResNet50 model consists of 50 layers, including convolutional layers, batch normalization layers, activation functions, and fully connected layers. .It is commonly used for image classification tasks, such as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC).

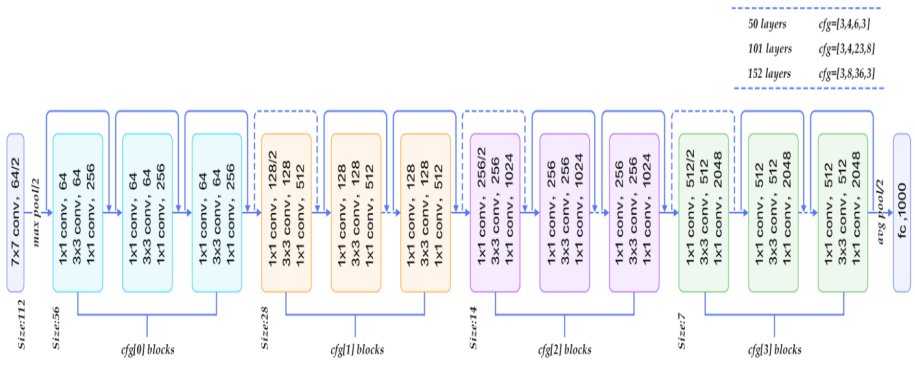

Figure 6: Illustration of Resnet50 model

Feature extracted from collection of images will be saved as Numpy file (.npy). When user chooses the query image, this query image will be extracted to vector and also save as Numpy file. Then system calculate the L2 distance (Euclidean Distance) from query vector to all vector of images in the collection. Our system choose 20 vectors from collection that nearest the query vector, then it shows top 20 corresponding images.

We also this system into web app using Flask [**7**], this is a micro web framework for Python, it is easy to use even for a newbie.

Our code is referenced from matsui528's Github [**6**].

4 Experiments

4.1 Some easy and hard cases

We can see in **Figure 7** and **Figure 8**, scene and object are too unique so that IR system can easily retrieve relevant images.

With a good viewpoint and good lighting condition, IR system will easily to retrieve relevant images, but We can see in **Figure 9** and **Figure 10**, with two query images of building that captured at night and gloaming. With the different lighting condition, IR system still working good if the building still can be saw.

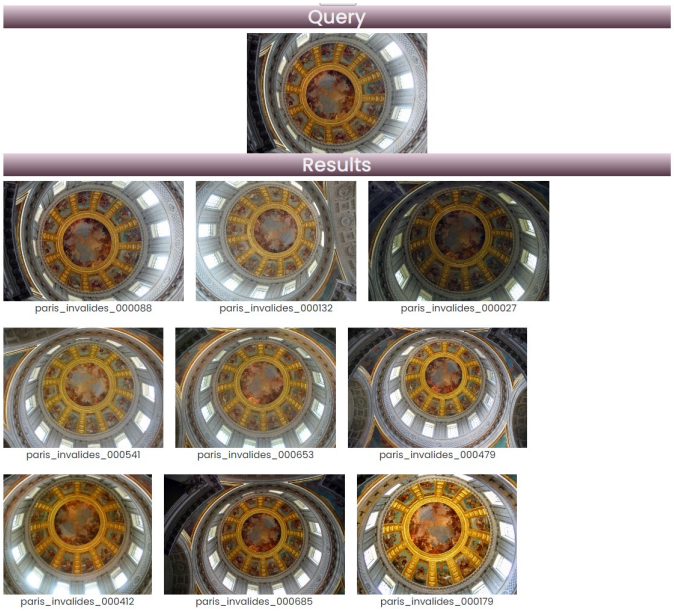

Figure 7: Illustration of easy case

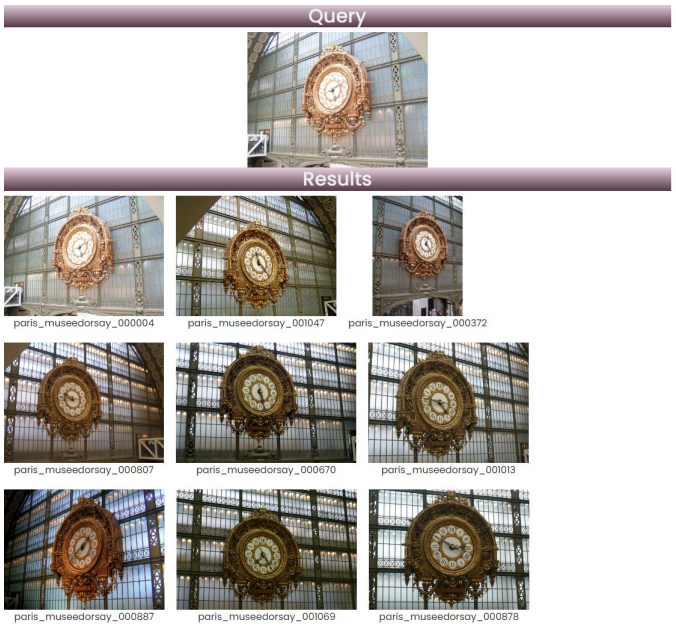

Figure 8: Illustration of easy case

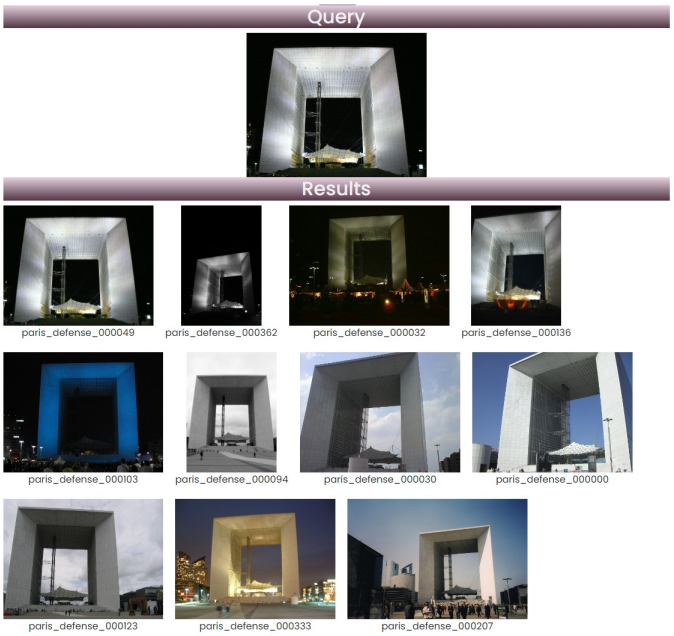

Figure 9: Illustration of easy case

Figure 10: Illustration of easy case

In **Figure 11**, we can see gray query image and the retrieval result is still good, so the color of image doesn't affect to the retrieval performance.

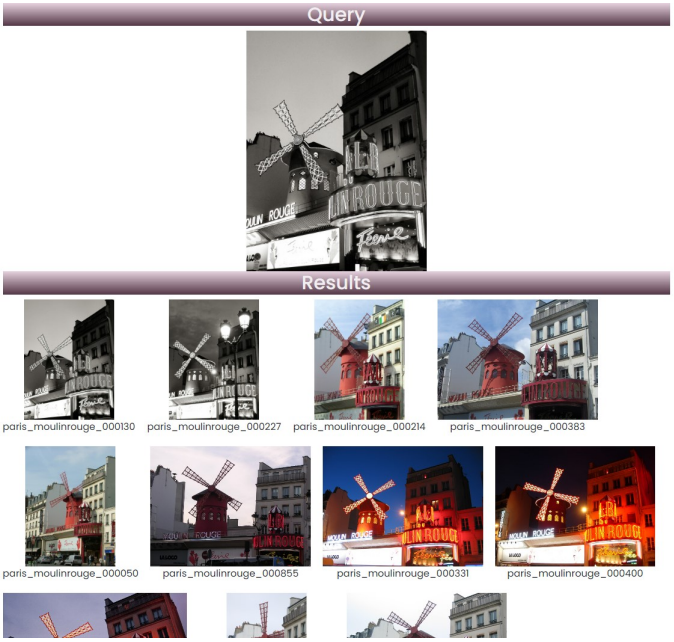

Figure 11: Illustration of easy case

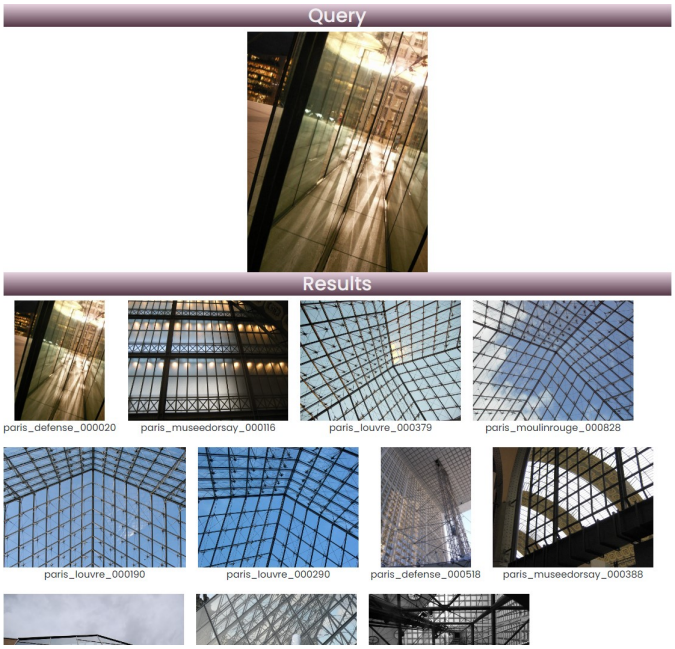

Figure 12: Illustration of hard case

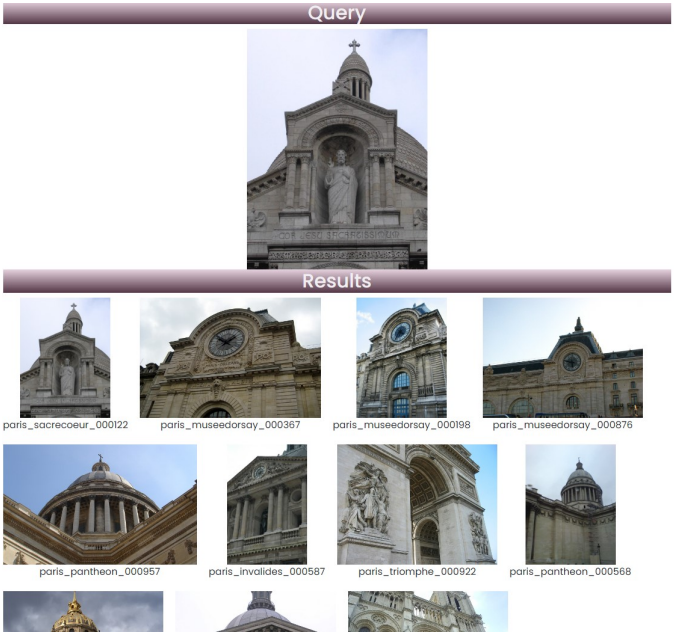

Figure 13: Illustration of hard case

In **Figure 12** and In **Figure 13**, IR system can only find one relevant image but that image is the same with the query image, we find the problem is the similarity image in same class of two above query are too few. We can solve this problem by add more images at several viewpoints and more hard cases.

In **Figure 14**, we can see in the query image, two women obscure the building lead to the difficulty in recognizing the building, it also make the IR system hard to know what is the thing that user want to find (human or building). All above reasons make the retrieval result are images of different building and images include human.

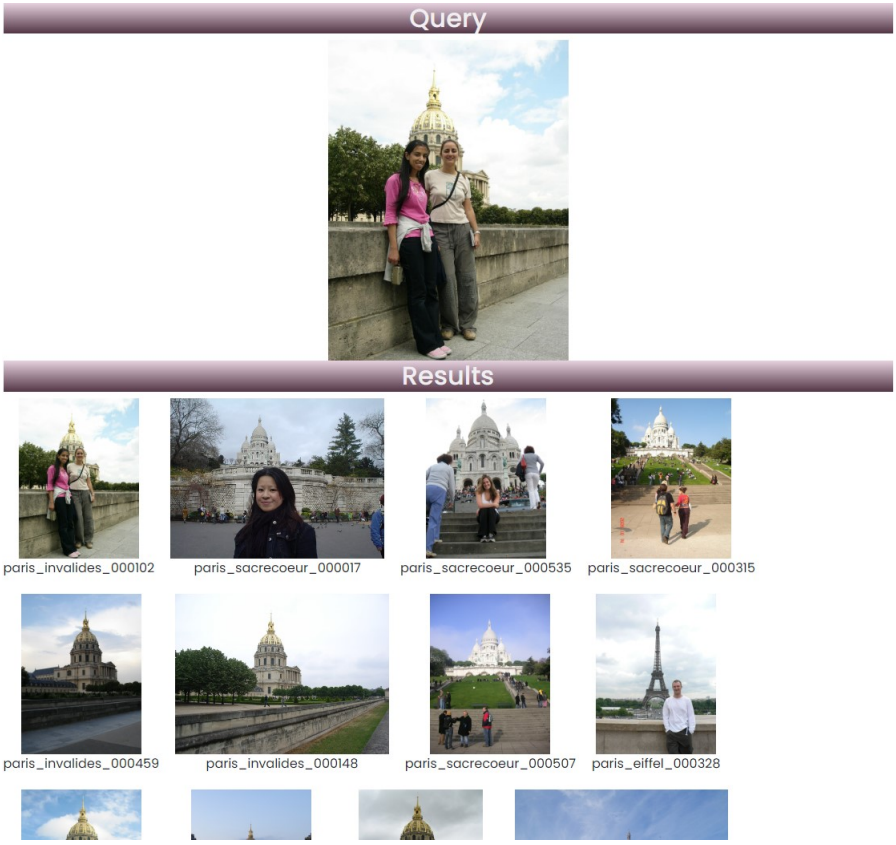

Figure 14: Illustration of hard case

One more hard case is the too far distance from the point of view to the building, we can see that in **Figure 15**, the building is too small, it also obscure by other buildings. It is easy to understand why IR system can't find relevant images, it is difficult even with human.

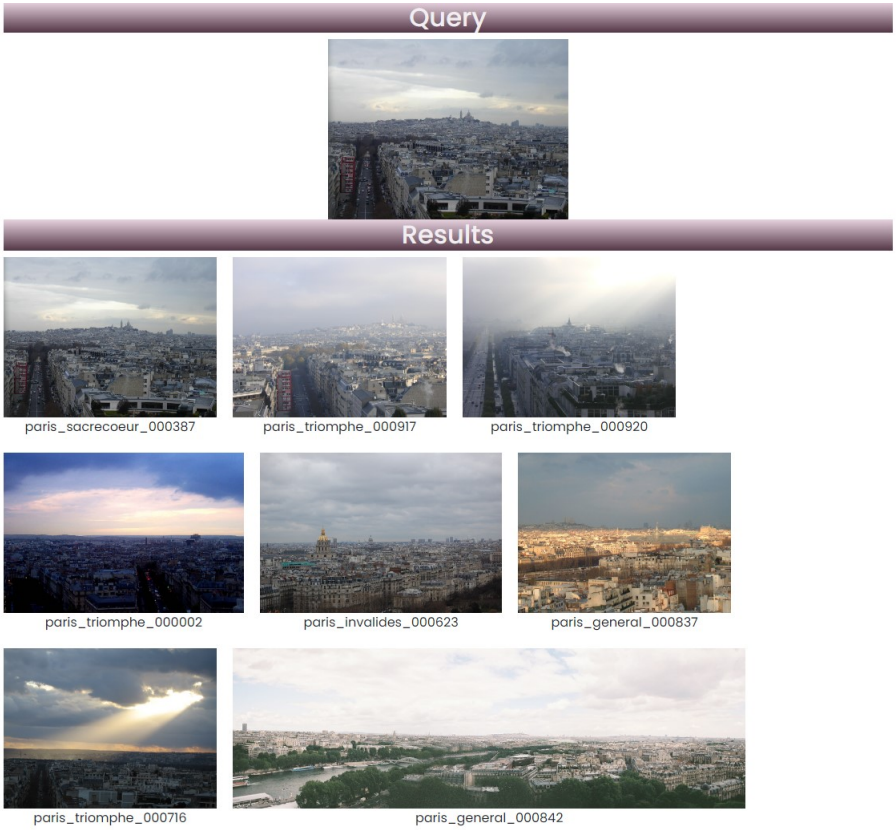

Figure 15: Illustration of hard case

4.2 Performance comparison

We use three networks in this project to have a comparison. Each network will be test with 55 query from 11 classes in Par6k dataset, we show top 21 similar images because top-1 retrieved image always is the same image with query image. We calculate AP for top 20 images retrieved (not include the query images). Then we calculate MAP (mean average precision) for each network.

Network	MAP
pretrained-VGG16	71.7
pretrained-VGG19	71.4
pretrained-Resnet50	75.2

Table 1: MAP of three network used in IR system

We extensively compared our results on each model in **Table 1**. We found the best result in model pretrained-Resnet50 network. Because the number of parameters of the model is complex and fit the large dataset so it gives more accurate results. Resnet50 also give a better result with hard case in **Figure 15**, the result is not too good but better than VGG16 and VGG19.

5 Conclusion

In this report, we introduced about Image Retrieval, build an IR system for Par6k and Oxford5k dataset, then we compared the performance of three networks. And from our experiments, we also pointed out some easy and hard cases, proposed some solution for hard cases.

References

[1] James Philbin, Relja Arandjelović and Andrew Zisserman(2018). The Oxford Buildings Dataset.Available at : https://www.robots.ox.ac.uk/ vgg/data/oxbuildings/

[2] The example images of the Paris dataset — download scientific diagram (no date). Available at: https://www.researchgate.net/figure/The-example-images-of-the-Paris-dataset_fig2_336888700.

[3] Khandelwal, V. (2020) The architecture and implementation of VGG-16, Medium. Towards AI. Available at: https://pub.towardsai.net/the-architecture-and-implementation-of-vgg-16-b050e5a5920b.

[4] Tomar, N. (2023) VGG19 UNET implementation in tensorflow, Idiot Developer. Available at: https://idiotdeveloper.com/vgg19-unet-implementation-in-tensorflow/

[5] Rastogi, A. (2022) RESNET50, Medium. Dev Genius. Available at: https://blog.devgenius.io/resnet50-6b42934db431

[6] matsui528 MATSUI528/SIS: Simple image search engine, GitHub. Available at: https://github.com/matsui528/sis.

[7] Welcome to Flask - Flask Documentation (2.2.x). Available at: https://flask.palletsprojects.com/en/2.2.x/.