Project 2 - Image Retrieval

Dat Nguyen Duy - 20520435[1], **Tung Le Xuan - 20520347**[2], **Kha Le Nhat - 20520208**[3]

[1,2, 3] Honor Program, Faculty of Computer Science, University of Information and Technology, Ho Chi Minh City, Vietnam

February 20, 2023

Abstract

The aim of every information retrieval (IR) system is to deliver relevant documents to an user's information need (IN). User's need is important to the quality of the system's search results. However, many IR systems ask the users to appreciate their information needs and interact them to the system, usually in form of queries. The systems assume the queries to be a perfect assessment of the information needs and relevant information, ending the interaction. However, in some cases database is very big, which challenges us to find out fast retrieval way.
.

Keywords: *Image Retrieval; Query; Faiss; Quantizaion*

1. Introduction

The image retrieval system is used for retrieving images related to the user request from the database. This is a computer system used for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, title or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web-based image annotation tools.

2. Dataset

The Paris Dataset includes 6412 images collected from Flickr by searching for particular Paris landmarks. The following 12 queries 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, Paris.

The Paris dataset consists of images provided by Flickr and use of these images must respect Flickr's Terms and Conditions of Use. In addition, use of the Paris dataset must follow our Terms of Access. For privacy issues with the dataset, please refer to our Dataset Privacy Notice.

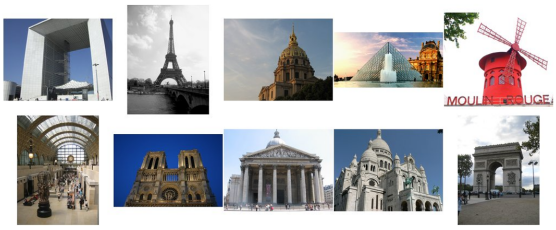

Figure 1. The Paris Dataset

The objective of our research is to improve better methods for searching for specific objects in extremely large datasets of images. In this work we focus on searching large collections of downloaded Flickr images - for Paris dataset.

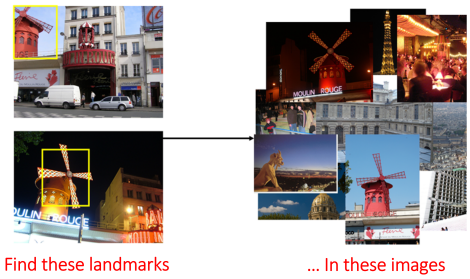

Figure 2. How to query and show result

3. Methods

Below, Figure 3 illustrates the architecture that demonstrates how to query to information retrieval system. Both offline and online are mentioned. For more details, we will explain feature extraction and indexing because they are two main steps which affect the performance of system. Especially, we use the state of the art model and faiss technique from Facebook to improve query system.

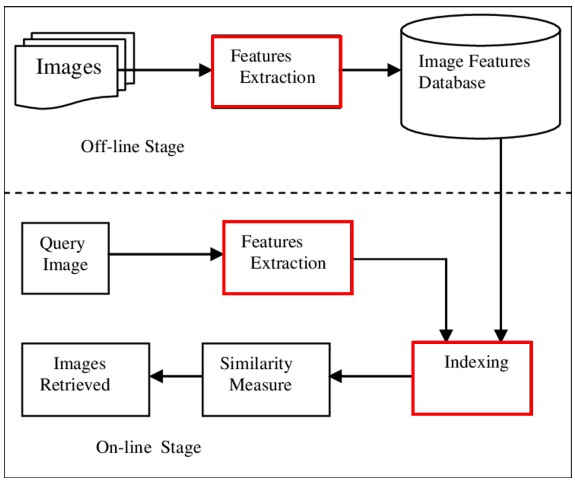

Figure 3. Offline and Online form - Query architecture

3.1. Feature Extraction

We also experiment with 4 methods: VGG16, ResNet50, RGBHistogram and Local Binary Patterns (LBP). The following explaination explicit the effectiveness of them.

3.1.1. VGG16

The benefit of using multiple smaller layers rather than a single large layer is that more non-linear activation layers accompany the convolution layers, improving the decision functions and allowing the network to converge quickly. That means decreasing memory problem.

Morever, VGG uses a smaller convolutional kernel, which reduces the network's ability to overfit during training exercises. A 3×3 filter is the optimal size because a smaller size cannot scan left-right and up-down information. Thus, VGG is the smallest possible model to understand an image's spatial features. Consistent 3×3 convolutions make the network simple to train.

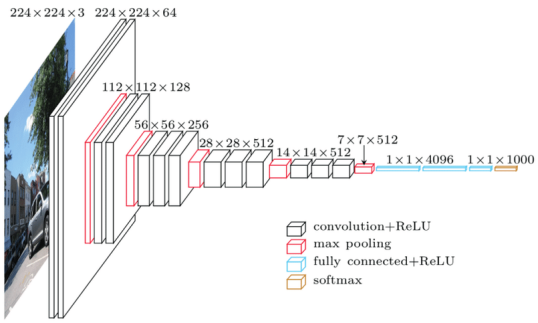

Figure 4. VGG16 architecture

3.1.2. ResNet50

ResNet50 is a 50-layer convolutional neural network (48 convolutional layers, one MaxPooling layer and one average pooling layer). Residual neural networks are a type of artificial neural network (ANN) that is designed with networks by stacking residual blocks.

The ResNet50 uses a bottleneck structure for a block. A bottleneck residual block uses 1×1 convolutions, known as a "bottleneck", which reduces the number of training parameters and matrix multiplications. This enables much faster training of each layer. It uses a stack of three layers rather than two layers

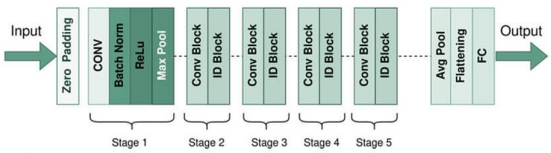

Figure 5. ResNet50 architecture

3.1.3. RGBHistogram

The main purpose of a histogram in computer vision is to illustrate us the distribution of certain values in the tones of image. The information displayed can help you determine the accuracy of the contrivance of image and adjust settings.

Histograms that relate with the specific color channels that are involved in create the colors of image are referred to as RGB histograms, which represent red, green and blue – the three basic colors used to generate all the colors that can be found in any image. Tones also refer to the different levels of brightness that can be found in image, and a luminance histogram shows the distribution of brightness throughout your photo.

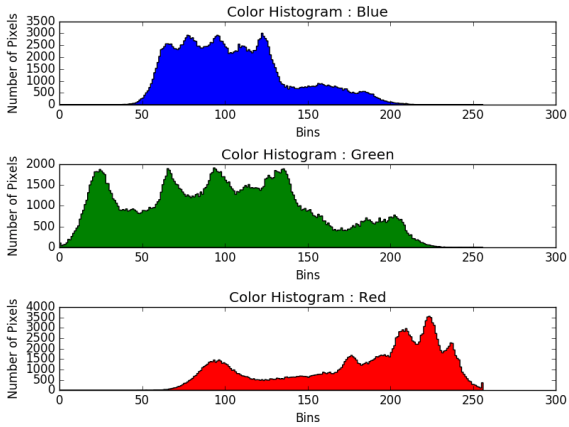

Figure 6. RGB Histogram

3.1.4. Local Binary Pattern

Local binary patterns (LBP) is a type of visual descriptor used for classification in computer vision. LBP has since been found to be a powerful feature for texture classification; it is combined with the Histogram of oriented gradients (HOG) descriptor, it improves the detection performance considerably.

LBP can be seen as a common approach to the traditionally different statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused.

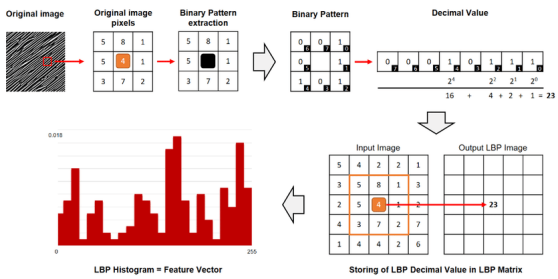

Figure 7. Local Binary Pattern

3.2. Indexing and similarity search

3.2.1. Introduction to Faiss

Our goal is a set of relevant image, which is indirectly implemented throughout calculating similarity measures. Similarity measure is based on indexing technique. Indexing is the way to get an unordered set of image into an order that will maximize the query's efficiency while searching or retrieval.

Faiss is a library for efficient similarity search and clustering of dense vectors. It consists of searching algorithms and implements in sets of vectors of any size, up to ones that possibly do not fit in RAM. It supports source code for evaluation and parameter tuning. Faiss includes the most useful algorithms are implemented on the GPU.

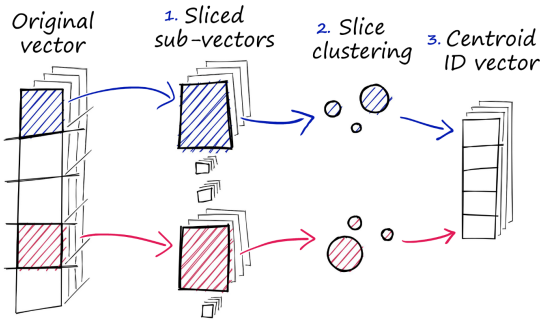

Figure 8. Facebook AI Search Similarity - Faiss

Faiss also can return not just the nearest neighbor, but also the 2nd nearest, 3rd, ..., k-th nearest neighbor. It searches several vectors at a time rather than one (multi-index or batch processing). For many index types, this is faster than searching sequentially.

There is a trade-off between speed and memory. If we accept wrong result 10 percentage of the time, we will have a method that is 10x faster or save 10x less memory. Faiss stores index on disk instead of RAM, especially index is formed binary vectors compared to floating-point vectors.

3.2.2. Main foundations used in Faiss

Inverted index: It is a data structure that searches similar object quickly. Faiss uses inverted index to store feature vectors and their position in feature space. This helps us to search nearest vectors fast.

Quantization: It is a generic method that relates to the compression of data into a smaller space. Quantization is used to reduce the memory footprint of indexes. This is an important task when comparing large arrays of feature vectors as they must all be loaded in memory to be compared. From that, we can improve the acceleration of retrieval.

Product Quantization: This part applies reducing dimensionality in feature space. Faiss use product quantization with the aim that feature vectors is divided smaller components or subvectors. After that, k-means clustering is used to find the centroids with components, respectively.These centroids with unique IDs and each ID represents a centroid. This helps us to reduce the size of dataset and speed up for searching.

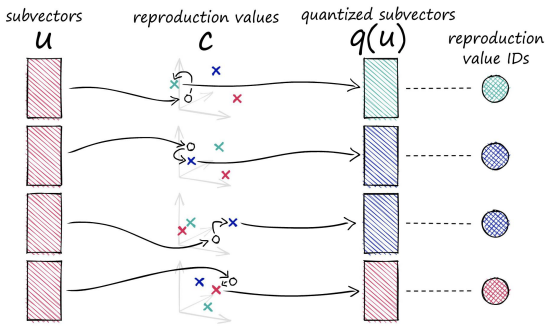

Figure 9. Quantization technique

4. Results

The below results or demo are implemented circumspectly. We use Streamlit framework to build image retrieval system. With the input query is a image or any cropped window of that image, our expectaion is a set of relevant image from available database.

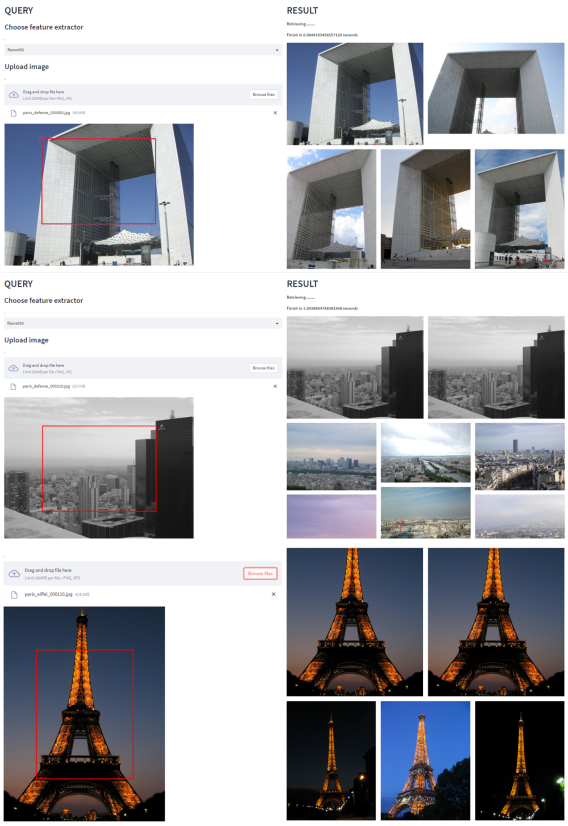

Figure 10. Results with system deployment

5. Conclusion

We have presented the goal of image retrieval, dataset contains 6K images and state of the art method is Faiss. Content is reported and results is demonstrated. We try to research both of them to expect as possibly as the best project. Although there are some remained problems that we can not generalize above, as soon as the we can get a better hardware in the future, the input data and the algorithms are better.

6. Acknowledgement

We thank to PhD. Thanh Ngo Duc from University of Information and Technology - Ho Chi Minh City who provided insight and improvement that greatly assisted the research.

References

[1] ttps://sites.google.com/uit.edu.vn/cs336project

[2] ttps://www.kaggle.com/datasets/skylord/oxbuildings

[3] ttps://engineering.fb.com/2017/03/29/data-infrastructure/faiss-a-library-for-efficient-similarity-search/

[4] ttps://github.com/facebookresearch/faiss/wiki

[5] ttps://www.pinecone.io/learn/product-quantization/

[6] ttps://datagen.tech/guides/computer-vision/vgg16/

[7] ttps://datagen.tech/guides/computer-vision/resnet-50/

[8] ttps://phlearn.com/magazine/histograms-better-photos/

[9] ttp://www.scholarpedia.org/article/LocalBinaryPatterns

[10] ttps://en.wikipedia.org/wiki/Imageretrieval