Content Based Image Retrieval

Zeynep Gokce
Department of Computer Engineering
Hacettepe University
Ankara,Turkey
n19248724@cs.hacettepe.edu.tr

***Abstract*—Within the advance of technologies, storage, and acquisition of information especially image data have come into prominence. Content-Based Image Retrieval System (CBIR) has an important role and aims to obtain the image(s) among a large collection of images which is similar to the query image given by the user. In this project, the retrieval system is proposed using distance-based k nearest neighbor approach. In order to make it faster, K-means clustering algorithm is utilized in the project. Experimental results on Inria Holidays database show promising performance.**

I. Introduction and Related Works

There have been lots of studies in the literature about image retrieval systems which are text-based and content-based approaches for image retrieval [1]. Content-Based Image Retrieval (CBIR) system is a kind of retrieval system that searches and retrieves similar images in a database using visual information such as color textures and etc. The images are represented by fixed-size vectors composed of hand-crafted features such as GIST, SIFT, FAST, and learning-based features such as representative activation values of any layer of deep convolutional deep neural networks for the representation of color, texture, structure, shape and etc [2].

CBIR system has two main processes. (1) feature extraction is the generation of representative descriptors for images. (2) image matching is to retrieve images in terms of the similarity values between query and images in the database. This process selects the best matches using the similarities.

In the literature, there has been significant progress in the image retrieval system. The main approaches can be divided into two groups which are (1) classical retrieval based on local features (bag-of-words, spatial verifications, Hamming embeddings) (2)deep learning based on global descriptors (CNN models). The deep convolutional networks are commonly proposed for feature extraction part [10]–[14]. CNN networks are used to learn the representative content features of the images. Deep CNN networks generate the high-dimensional features and outperform the hand-generated image feature-based retrieval systems. One of the popular image retrieval systems is DELF [15] network which has combined local features with deep learning-based features. It is based on the nearest neighbor method. In order to enhance the model result, the verification part is proposed to match the features. Another popular and powerful image retrieval system is proposed by Andrei Boiarov et.al [16]. The feature extraction is done by ResNet50. The similarity value between these features is calculated by dot product. Images are retrieved using threshold value on dot product similarity values.

II. Methodology

In this project, pre-trained ResNet50 is employed to extract the representative features contains the contents of the image such as texture, shapes, and colors. In essence, the images in the database will be indexed with visual contents. In order to retrieve images that are relevant to the query image, the closest images are obtained using the distance function as a measurement of relevance. The distance metric in K-NN approach is proposed to get the relevant image to query. To retrieve the related images faster K-means clustering algorithm is proposed. The methodology steps are shown in Figure 1 and the detail of the proposed methodology has explained in three main steps in the following sections.

A. Feature Extraction

Feature extraction is the processing for the representation of the images on the database for image retrieval systems. As a general term, the features are divided into two different types which are text-based and visual-based features. In this project, visual features are utilized for the image retrieval system. Visual features contain the content-based (such as color, texture, and edge features) information about the images. Instead of classical feature extraction methods, deep convolutional features from deep models (such as VGGNet19, ResNet50, InceptionV3 which are trained for classification) are proposed for the image retrieval system. The features are extracted from the last layers of the deep pre-trained convolutional network for all images in the database and also for the query. In essence, the images in the database will be indexed with visual contents using these deep convolutional features.

In this project, the ResNet50 [7] trained on ImageNet [5] dataset is used and the features of images are extracted from the last average pooling layer of the network. The images are resized to 224x224 pixels before given to the network. Each image is represented by 1x2048 dimensional feature vector.

B. Measurement of Relevance

The aim of Content-Based Image Retrieval Systems is to retrieve the images from the image database that are similar/relevant to the query image.

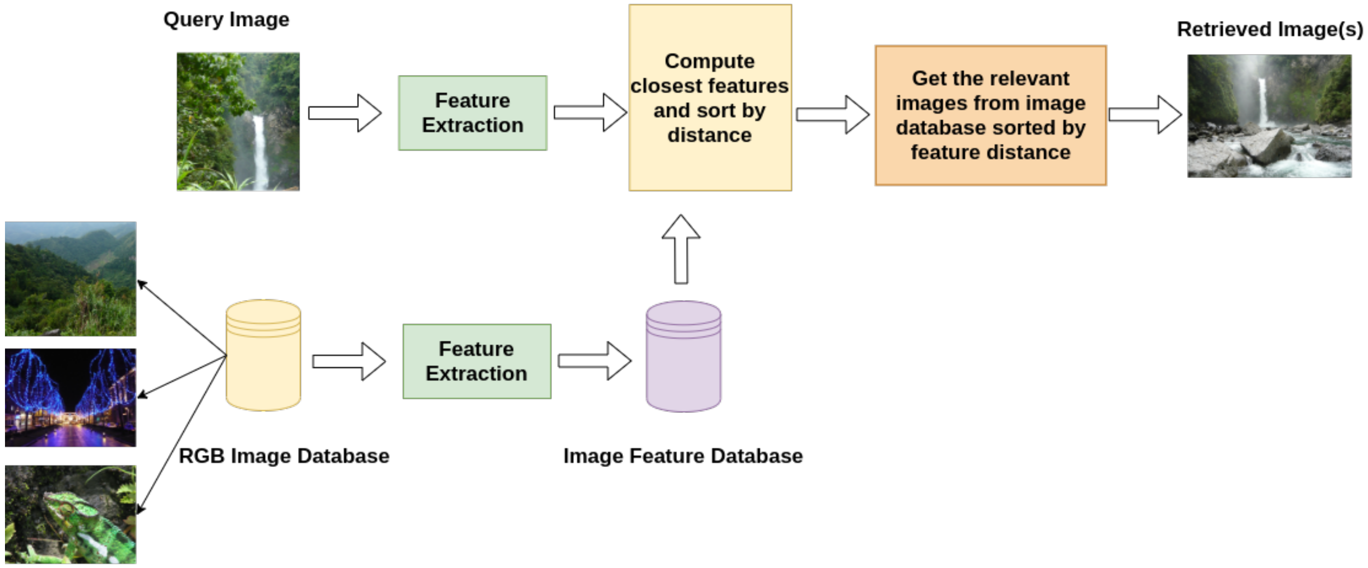
Query Image
Feature Extraction
Compute closest features and sort by distance
Get the relevant images from image database sorted by feature distance
Retrieved Image(s)
RGB Image Database
Feature Extraction
Image Feature Database

Fig. 1: Content Based Image Retrieval Model

In order to retrieve images that are relevant to the query image, the similarity measurement criteria are performed. The closest/similar images are obtained using the distances between the query and image database. Similarity using the distance between images should be measured in the feature space. The various distance functions are applied for the similarity evaluation.

The euclidean distance function is the most widely used similarity metric. It can be described as the square root of the sum of the squares of the differences between vector components. The Euclidean distance between feature vectors I and Q can be defined as

$$D = \sqrt{\sum_{k=1}^{n} (I_k - Q_k)^2}$$

where the n is the length of the feature vector.

There are also better distance measures for high dimensional features such as Manhattan distance metric. It is expressed by the following formula

$$D = \sum_{k=1}^{n} |I_k - Q_k|$$

C. Retrieving the Image

Using the distance metrics, the similarity can be measured between two images. The smaller the distance between the two images, the more similar these two images are. In order to decide the most relevant images from the database, top k nearest/closest images to the query image are retrieved using distance functions. The simplest K-NN algorithm, distance-based algorithm. It is needed to calculate all distances between the query and each image in the database. Finally, it finds the K closest (most similar) images to the query.

In the retrieval part, K-NN algorithm is a simple distance-based brute-force implementation. In order to make retrieving relevant images faster using distance function, the images in the database are clustered using K-Means algorithm firstly. Instead of looking at all images in the huge database, the distance function is applied firstly on the centroids of the clusters to find the relevant cluster. Then, the top k relevant images are retrieved in the relevant cluster images by ranking. As a novelty, the k-means algorithm is proposed for images in the database to make the retrieving part faster compared to a similar approach [3].

III. Experiments & Discussion

This section contains some implementation details and discussion about the results. The related images to the query image is retrieved from the database in two steps (1) using k nearest neighbor approach (2) using k nearest neighbor approach based on clusters trained with K-Means algorithm. Both approaches are implemented, evaluated with mAP metric, and explained in detail. For the first approach (1), one more metric (label ranking average precision score) is applied for evaluation of the model except for mAP.

A. Dataset

INRIA Holidays Dataset [9] consists of 1491 high-resolution images that are taken in different places from a variety of scene types (natural, water, fire effects, etc). These images have a different translation, rotation, or viewpoints. There are 500 query images in this dataset. According to the dataset, there are 1-10 relevant images for each query image. This is evaluated using mean average precision (mAP) defined in [9].

B. Image retrieval with K-Nearest Neighbor

The first experiments are performed by retrieving top k similar images using only k nearest neighbor approach. The simi-

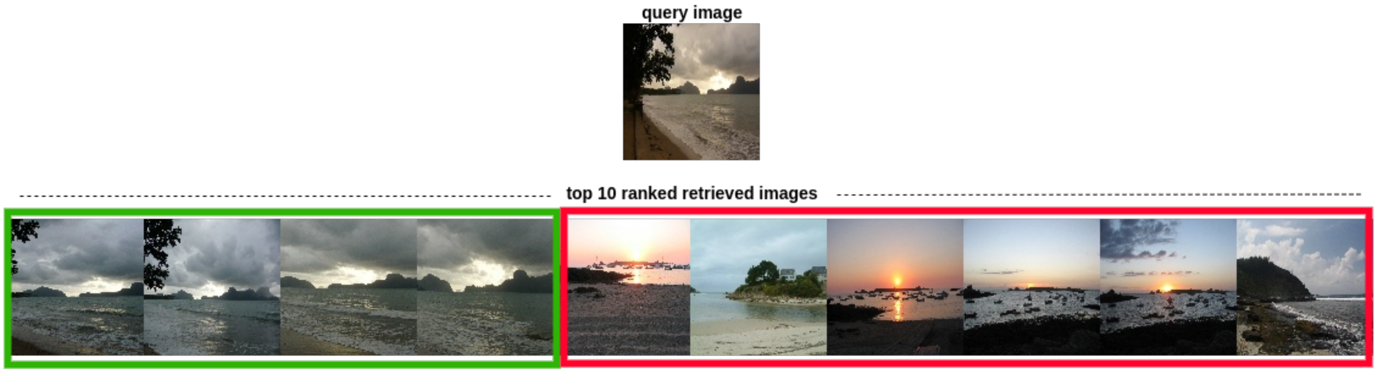

Fig. 2: Distance based k nearest neighbor image retrieval system results is given with the top 10 images. The images covered with green color are the ground truth images while the other images covered with the red color are the retrieved images from the other classes.

	mAP@1	mAP@5	mAP@10
VGG19 [6] - Euclidean	0.49	0.64	0.66
VGG19 [6] - Manhattan	0.50	0.67	0.68
ResNet50 [7] - Euclidean	0.55	0.74	**0.75**
ResNet50 [7] - Manhattan	0.53	0.71	0.72
InceptionV3 [8] - Euclidean	0.49	0.67	0.69
InceptionV3 [8] - Manhattan	0.50	0.67	0.69

TABLE I: Evaluation results (mAP) for retrieved top k im that are relevant.

larity is obtained using different distance functions (Euclic and Manhattan) and different features from various deep volutional networks (VGG19, ResNet50, and Inception According to Table I, the best performance (using mAP me is achieved with ResNet50 network and Euclidean distance at top 10 relevant images. Euclidean distance metric gives better results in ResNet50 features while Manhattan distance metric gives better in VGG19 features. It is clear that the ResNet50 features are more comparable and representative in distance-based comparison. Although this k nearest neighbor is a simple approach, it gives an effective result.

Another evaluation metric, *label ranking average precision score* (rank and label based, scikit learn function), is used for the evaluation of the nearest neighbor approach. Due to the fact that each query image has a different number of the relevant images in the database, this metric is more proper for these kinds of systems taking into account both rank and label. According to the given Figure 3, it is clear to say that the model score decreases while the number of retrieving images increases. The qualitative result is shown in Figure 2. Although the retrieved images come from different labels, al retrieval images are consistent.

C. Image retrieval with K-Means

K-NN algorithm is a simple brute-force implementation. In order to retrieve top k relevant images, the distance between all images in the database and the query should be calculated. It was not an efficient and not user-friendly implementation. To retrieve images in an efficient way, K-means algorithm is implemented to cluster the image database. The aim is to cluster the images with related scenes together.

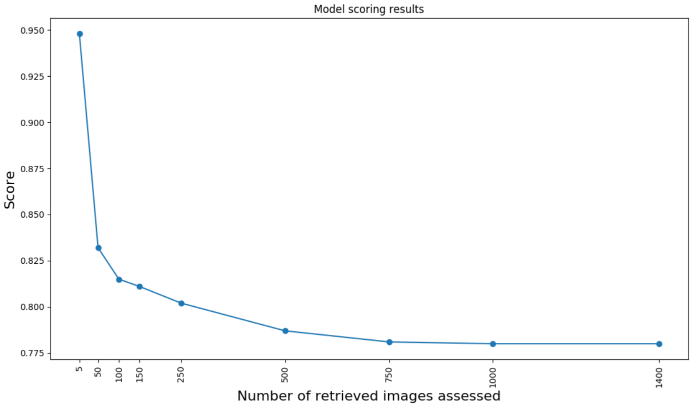

Fig. 3: Ranking Average Precision Score of model with different number of retrieved images.

As mentioned in the experiments on Table I, Euclidean distance, and ResNet50 features are utilized for the next experiments with K-means. The database is trained with different numbers of clusters(5-500), the higher performance of the image retrieval system is achieved **0.70 mAP** value with 5 cluster centers based on top 10 results.

The main challenge is to define the cluster centers in the dataset with features. The dataset groups are so similar to each other. If the cluster centers could not be obtained properly, the retrieving part returns the wrong result. For this reason, the performance of the model decreased to 0.61 mAP as the number of clusters increased.

On the other hand, in the first experiment (without K-means), the best-retrieved results(see Table I) are obtained in 209 seconds while the images using K-means are retrieved in 2.79 seconds with 200 cluster centers. Both experiment do not give the same performance but the proposed model(k-means) is 70x faster than the model(without k-means).

IV. Conclusion

In this project, a faster and acceptable image retrieval system is proposed. Each image is represented as a feature vector by using ResNet50 fully connected layers. The related images to the query are retrieved using the Euclidean distance function. With using distance-based k nearest neighbor approach, experimental results show that the image retrieval system achieved high performance. In order to make the system faster, the k-means clustering applied. All experiments are evaluated on Inria Holidays Database.

References

[1] Alemu, Yihun, et al. ”Image retrieval in multimedia databases: A survey.” 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE, 2009.

[2] Zhou, Wengang, Houqiang Li, and Qi Tian. ”Recent advance in content-based image retrieval: A literature survey.” arXiv preprint arXiv:1706.06064 (2017).

[3] Seddati, Omar, et al. ”Towards good practices for image retrieval based on CNN features.” Proceedings of the IEEE International Conference on Computer Vision Workshops. 2017.

[4] Jgou, H., M. Douze, and C. Schmid. Hamming Embedding and Weak Geometry Consistency for Large Scale Image Search-extended version. Research Report 6709, Oct. 2008. 2.

[5] ImageNet. http://www.image-net.org

[6] Simonyan, Karen, and Andrew Zisserman. ”Very deep convolutional networks for large-scale image recognition.” arXiv preprint arXiv:1409.1556 (2014).

[7] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770778, 2016.

[8] Christian Szegedy, Vincent Vanhoucke, et aI, Rethinking the Inception Architecture for Computer Vision. arXiv: 1512.00567, 20 15.

[9] Jegou, Herve, Matthijs Douze, and Cordelia Schmid. ”Hamming embedding and weak geometric consistency for large scale image search.” European conference on computer vision. Springer, Berlin, Heidelberg, 2008.

[10] Tolias, Giorgos, Yannis Avrithis, and Herv Jgou. ”To aggregate or not to aggregate: Selective match kernels for image search.” Proceedings of the IEEE International Conference on Computer Vision. 2013.

[11] Wan, Ji, et al. ”Deep learning for content-based image retrieval: A comprehensive study.” Proceedings of the 22nd ACM international conference on Multimedia. 2014.

[12] Krizhevsky, Alex, and Geoffrey E. Hinton. ”Using very deep autoencoders for content-based image retrieval.” ESANN. Vol. 1. 2011.

[13] Gordo A, Almazan J, Revaud J, Larlus D (2016) Deep image retrieval: learning global representations for image search, in European conference on computer vision (ECCV)

[14] Paulin M, Douze M, Harchaoui Z, Mairal J, Perronin F, Schmid C (2015) Local convolutional features with unsupervised training for image retrieval, in IEEE International Conference onComputer Vision (ICCV), 9199

[15] Filip Radenovi, Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondej Chum. 2018. Revisiting Oxford and Paris: Large-Scale Image Retrieval Benchmarking.arXiv preprint arXiv:1803.11285 (2018)

[16] Boiarov, Andrei, and Eduard Tyantov. ”Large scale landmark recognition via deep metric learning.” Proceedings of the 28th ACM International Conference on Information and Knowledge Management. 2019.