



ENHANCING SKETCH-BASED IMAGE RETRIEVAL BY CNN SEMANTIC RE-RANKING

Anisha R, PG Student, Department of ECE, SA Engineering College, Chennai, India
N. Anusha, Professor, Department of ECE, SA Engineering College, Chennai, India
G.Kavya, Professor, Department of ECE, SA Engineering College, Chennai, India

ABSTRACT – A re-ranking system to enhance the performance of sketch-predicated image retrieval (SBIR). From the subsisting approaches, the proposed system can leverage category information brought by CNNs to fortify homogeneous attribute quantification between the images. To achieve efficacious relegation, one CNN model is trained for relegation of sketches, another for that of natural images. By training dual CNN models, the semantic information of both the images is captured by deep learning. To quantify the category homogeneous attribute between images, a category homogeneous attribute quantification method are proposed. Category informations are then utilized for re-ranking. Re-ranking operation first infers the retrieval category of the query sketch and utilizes the category kindred attribute quantification to quantify the category homogeneous attribute between the query and each initial retrieval result. Determinately, initial retrieval results are re-ranked.

KEYWORDS: convolutional neural network, re-ranking.

I. INTRODUCTION

Text-based image retrieval systems, retrieve images by some keywords. Although they are widely applied, keywords are sometimes not sufficient to express the desired pictures clearly. So, content-based image retrieval (CBIR) systems which retrieve images by exemplar images, emerge. Besides, algorithms for content-based video retrieval and processing exist. However, sometimes there is not an exemplar picture as the query image. At this very moment, the sketch-based image retrieval (SBIR) system is useful. SBIR systems just need users to draw a simple sketch with a few lines or shapes as the query image. Lines and shapes tend to reflect primary outlines of a desired object and bring little redundant information. Therefore, sometimes sketches are more capable of expressing users' search intentions. An effective and timesaving SBIR re-ranking method can significantly improve the performance of an SBIR system without adding much time cost.

II. CONVOLUTIONAL NEURAL NETWORK

In deep learning, there are different types of models such as Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN) and Reinforcement Learning. But there is one particular model that has been contributing a lot in the field of computer vision and image analysis which is convolutional neural network (CNN) or ConvNets. CNNs are a class of Deep Neural Networks that can recognize and relegate particular features from images and are widely utilized for analyzing visual images. Their applications range from image and video apperception, image relegation, medical image analysis, computer vision and natural language processing.

2.1 Architecture of CNN

There are two main components to a CNN architecture.

- A convolution tool that isolates and identifies the varied features of the image for analysis in a process called as Feature Extraction.
- A completely connected layer that utilizes the output from the convolution process and presages the class of the image predicated on the features extracted in previous stages.

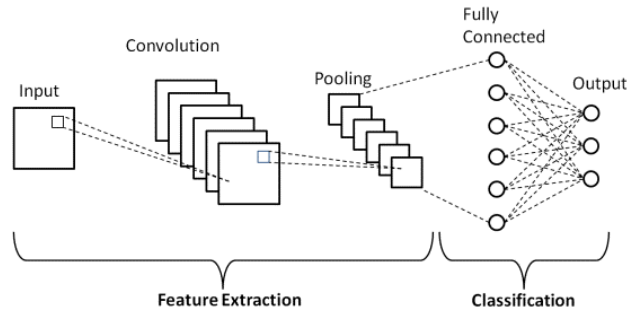


Figure: 1

2.2 Convolutional Layer

This layer is the first layer that is utilized to extract the varied features from the input images. In this first layer, the mathematical operation of convolution is executed between the input image and a filter of a particular size $M \times M$. By sliding the filter over the input image, the dot product is taken between the filter and the components of the input image with veneration to the size of the filter ($M \times M$).

The resultant output is defined as the feature map which gives details about the image such as the corners and edges. Later, this feature map is alimented to other layers to learn several other features of the input image.

2.3 Pooling Layer

Mostly, a Convolutional Layer is followed by a Pooling Layer. The primary aim of this layer is to decrement the size of the convolved feature map to truncate the computational costs. This is performed by decrementing the connections between layers and independently operates on each feature map. Depending upon method utilized, there are several types of Pooling operations.

In Max Pooling, the most astronomically immense element is taken from feature map. Average Pooling is calculated as the average of the elements in a predefined sized Image section. The total sum of the elements in the predefined section is calculated in Sum Pooling. The Pooling Layer conventionally accommodates as a bridge between the Convolutional Layer and the FC Layer.

2.4 Fully Connected Layer

The fully connected (FC) layer consists of the weights and biases along with the neurons and is utilized to connect the neurons between two different layers. These layers are normally placed afore the output layer and form the last few layers of a CNN Architecture.

In this, the input image from the layers are flattened and victualed to the FC layer. The flattened vector then undergoes few more FC layers where the mathematical functions operations normally take place. In this stage, the relegation process commences to take place.

III. WORKING AND ANALYSIS

Framework of the proposed SBIR re-ranking system is given in the following figure. Query refers to a query sketch, Image Dataset contains all the natural images that are needed for the paper. The query sketch and natural images are put into the SBIR system to engender the initial retrieval results. Q-Net and N-Net are two CNNs. Once they are trained, they are utilized to get the category information of the query sketch and the initial retrieval results, respectively. Determinately, with the avail of category information, initial retrieval results are re-ranked in re-ranking.

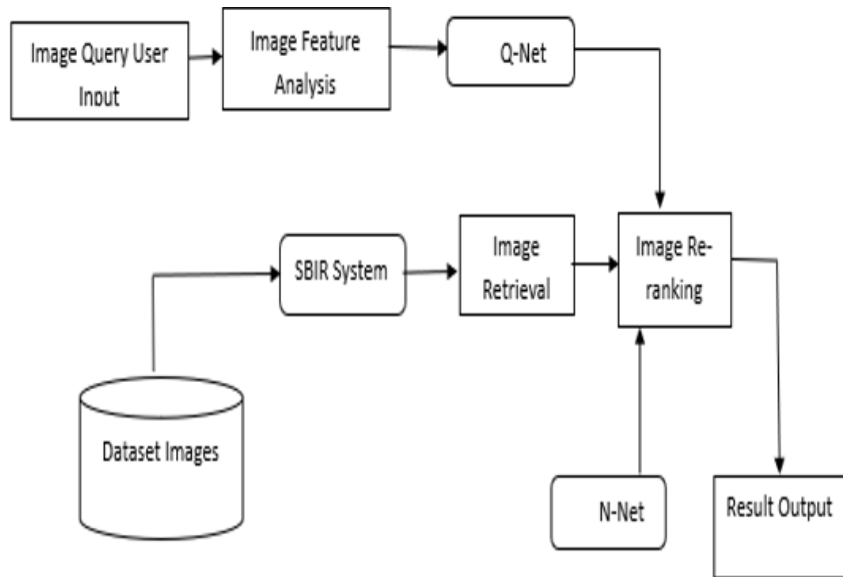


Figure: 2

In this paper, we test the performance of VGGNet, AlexNet, and GoogLeNet. The retrieved images and the comparison of accuracy and elapsed time are given below.

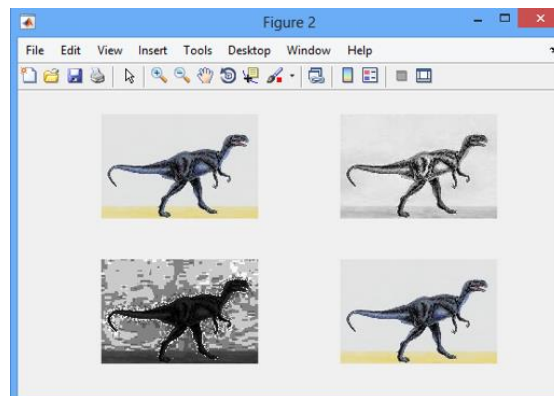


Figure: 3

The above figure 3 is the image retrieved from VGG CNN Architecture.

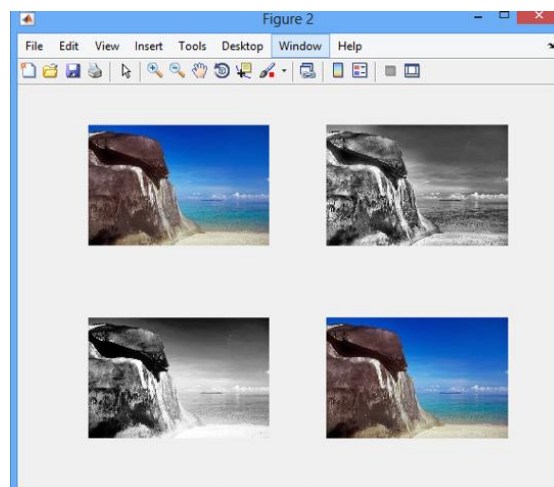


Figure: 4

The above figure 4 is the image retrieved from ALEXNET CNN Architecture.

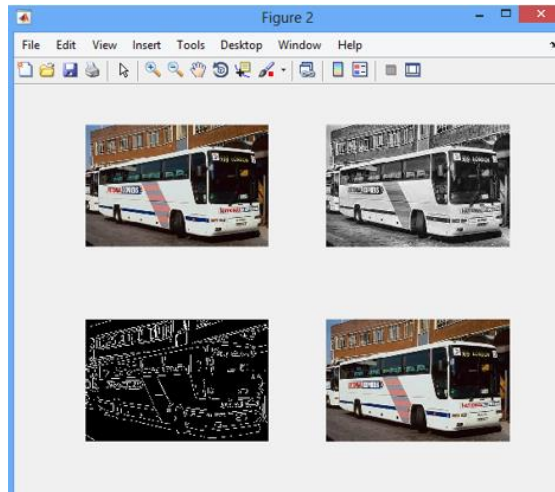


Figure: 5

The above figure 4 is the image retrieved from GOOGLNET CNN Architecture.

All the architectures retrieve images but the accuracy and the elapsed time differs which is compared in the following table. The image retrievals are done using MATLAB software and the outputs are received using a receiver unit which is a personal computer.

ARCHITECTURE	ACCURACY	ELAPSED TIME
VGGNET	57.4%	26 sec
ALEXNET	28.57%	24 sec
GOOGLNET	28.57%	4 min 28 sec

Table: 1

IV. RESULTS AND CONCLUSIONS

We propose a re-ranking-based SBIR system to enhance the performance of SBIR systems. First, we train two CNNs separately, where one is for sketch classification, and the other is for natural image classification. By this means, CNN models study the semantic information of sketches and natural images. After this, CNN-based image classification is carried out on a sketch and its initial retrieval results, and category information of sketches and natural images are obtained. Finally, the initial retrieval results are re-ranked through measuring the similarity between the category information of the query sketch and the initial retrieval results. Experiments show that our proposed re-ranking-based SBIR system significantly improves the performance of various SBIR systems. Thus the performance of different architectures for image retrieval has done successfully.

REFERENCES

- [1] A. Chalechale, G. Naghdy, and A. Mertins, "Edge image description using angular radial partitioning," *IEEE Proceedings-Vision, Image and Signal Processing*, vol. 151(2): 93–101, April, 2004.
- [2] Y. Cao, C. Wang, L. Zhang, L. Zhang. "Edgel Index for Large-Scale Sketch-based Image Search," *CVPR*, IEEE Conference, 2011 R. Nicole, "Title of paper with only first word capitalized," *J. Name Stand. Abbrev.*, in press.
- [3]] E. D. Sciascio, G. Mingolla, M. Mongiello, "Content-based image retrieval over the web using query by sketch and relevance feedback," *VISUAL'99*, London, UK, 1999, pp. 123–130.

- [4] C. Liu, D. Wang, X. Liu, C. Wang, L. Zhang, B. Zhang, "Robust semantic sketch based specific image retrieval". ICME, 2010 IEEE.
- [5] R. Datta, D. Joshi, J. Li, and J. Wang. "Image retrieval: Ideas, influences, and trends of the new age". ACM, Computing Surveys, 2008.
- [6] G. Salton and C. Buckley. "Improving retrieval performance by relevance feedback". Journal of the American Society for Information Science, 41(4): 288– 297, 1999
- [7] I. J. Cox, M. L. Miller, T. P. Minka, T. V. Papathomas, P. N. Yianilos, "The Bayesian Image Retrieval System, PicHunter: Theory, Implementation and Psychological Experiments", IEEE Transactions on Image Processing, 9(1), pp. 20–37, 2000.
- [8] E. Cheng, F. Jing and L. Zhang, "A unified relevance feedback framework for web image retrieval", IEEE Trans. Image Process., vol. 18, no. 6, pp.1350–1357, 2009.
- [9] P. Salembier, F. Marqués, Region-based representations of image and video: segmentation tools for multimedia services. Circuits and Systems for Video Technology, IEEE Transactions on, 1999, 9(8): 1147– 1169.
- [10] K. Hirata and T. Kato, "Query by visual example - content based image retrieval," in Proc. Adv. Database Technol. 1992, pp. 56–71.