CONTENT BASED IMAGE RETRIEVAL USING AUTOENCODERS

Project Report

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF TECHNOLOGY in

COMPUTER SCIENCE

by

Ayushi Jain (202CS003)

(Reg no. 202639)

DEPARTMENT OF COMPUTER SCIENCE

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALORE-575025

May, 2021

DECLARATION

I hereby declare that the Report of the P.G. Project Work entitled **CONTENT BASED IMAGE RETRIEVAL USING AUTOENCODERS** which is being submitted to the National Institute of Technology Karnataka Surathkal, in partial fulfillment of the requirements for the award of the Degree of Master of Technology (M.tech) in Computer Science Engineering in the Department of Computer Science, is a bonafide report of the work carried out by me. The material contained in this report has not been submitted to any University or Institution for the award of any degree.

Ayushi Jain (202CS003)

Registration number: 202639

Department of Computer Science

Place: NITK, SURATHKAL

Date: 15-05-2021

CERTIFICATE

This is to certify that the P.G. Project Work Report entitled **CONTENT BASED IMAGE RETRIEVAL USING AUTOENCODERS** submitted by ***Ayushi Jain (Register Number: 202639, Roll Number: 202CS003)*** as the record of the work carried out by her, is accepted as the P.G. Project Work Report submission in partial fulfillment of the requirements for the award of the degree of ***Master of Technology (M.tech)*** in Computer Science Engineering in the Department of ***Computer Science***.

Internal Guide
Dr. Vani M
(Associate Professor)

Chairman - DPGC
(Signature with Date and Seal)

ACKNOWLEDGEMENT

I would like to take this opportunity to express my gratitude to all those who have contributed to the completion of the mini project and helped me with valuable suggestions for improvement. I am extremely grateful to **Shashidhar G Koolagudi (HOD),** Department of Computer Science and Engineering, for providing me with the best facilities and atmosphere for creative work guidance and encouragement. I would like to thank my coordinator, **Mrs. Vani M,** Department of Computer Science and Engineering, Surathkal for all help and support extended to me. I thank all Staff members of my college and friends for extending their cooperation during my project.

Above all I would like to thank my parents without whose blessings, I would not be able to accomplish my goal.

Ayushi Jain

CSE MTech(202CS003)

NITK SURATHKAL

ABSTRACT

With the advancement in technology, the amount of digital media content has seen an exponential growth. Internet and storage devices are filled with huge amounts of useful digital media content. But, unlike textual data, visual data cannot be indexed making it much more difficult to extract the images of our interest. Content based image retrieval is a technique for retrieving relevant content from a collection of images based on visual information.

This project focuses on using autoencoders for content based image retrieval. Autoencoders are a different type of neural network that learn a compressed form of input data. Encoders convert the images into a compressed latent space representation(LSR), which are then stored. When a query image is input it's LSR is compared with LSR of other images and similar images are found.

Table of Content

1. Introduction

Content based image retrieval is the task of finding images that are similar to a query image based on content from the database.

Web contains a huge amount of information. So, we must be able to explore it in an efficient way and make the most of it. Due to advancement in technologies we are creating huge amounts of visual data. The amount of visual data is much more than textual data but image data cannot be indexed like textual data which makes its retrieval much more difficult. Therefore a Content based image retrieval has become a good area of research.

Suppose we find something appealing somewhere, like an outfit or a pet or any flower. We may like to find more stuff similar to it on the internet. Many search engines are looking forward to providing an image based search. Wherein apart from text we can also upload an image query and get results based on it.

While browsing through an online shopping website we may like to see more products similar to the products we like. CBIR provides a way to improve recommendation systems in these websites. This will enhance the browsing experience of a person by showing them what they like to see. CBIR can also be used to provide a search based on images where the user can upload a pic of the product and get similar products as suggestions.

Many websites (visual discovery engines) like at Pinterest, StitchFix, and Flickr—started using Deep Learning to learn representations of their images, and provide recommendations based on the content users find visually pleasing. It helps them understand what a user is interested in and show them content according to what they like.

CBIR is also used for face recognition. This allows consumers to search photo libraries efficiently. Consumers may like to retrieve photos of a particular person from the library, this can be done by CBIR.

As digitization is becoming popular in medical areas. More and more organizations are shifting towards storing images from MRIs, X-rays in digitized format. Content based image retrieval plays an important role in mining similar images from the database.Clinical decision support techniques such as case-based reasoning or evidence-based medicine can produce a strong need to retrieve images valuable for supporting certain diagnoses. For the clinical decision-making process, it can be beneficial or even important to find other images of the same modality, the same anatomic region, or the same disease

Teaching and research in medical can benefit widely from studying and analyzing visual information present in databases. Content Based image retrieval can aid the process enormously.

Problem Definition:

The aim of the project is to develop an image based search engine. Image-based queries will be passed to the system and the system will retrieve images that are visually similar to the given image. The system will work similar to text based search engines but instead of text image based query will be passed.

The main idea is to use a different type of neural network called autoencoders which is an unsupervised learning algorithm. Autoencoders will convert an input image into a compressed Latent Space Representation (LSR).When an image is searched, the LSR of the input image is taken and compared with the feature vectors of other images in the database. Only those images from the database are returned where the distance between LSR of input and other images is considerably less.

2. Literature Survey

Content based image retrieval is the task of retrieving images similar to a query image. CBIR can be carried out based on various parameters like shape, colour, texture etc. There are two steps in CBIR feature extraction and feature matching. Feature extraction is the main step. Once the suitable features are extracted they can be matched using different parameters like distance measures.

In this section recent trends on content based image retrieval have been explored.

2.1 Texture based CBIR

The texture of an image shows how smooth, rough or regular it is. There are three methods generally used in image processing the region's structure, construction statistics, and spectra a. The texture and homogeneity of an image is a property that is not affected by intensity or if there is only a single colour.

[1] proposed a correlation model for the extraction of features of an image. The method is based on Gray-Level Co-Occurrence Matrix (GLCM). Each image in the data set is converted into gray level image and then Gaussian Mixture Model (GMM) is applied. The features are extracted from GLCM and are given as input to the model-based technique so that the relative Probability Density Functions (PDF) are extracted. The Kullback-Leibler divergence method (KL-Divergence) is used to compare the relative probability density functions of training and test data.

[2] proposed a method where SIFT and TURF technologies are combined. It improves the image recovery process by allowing extraction of image features and their matching together.

[3] proposed PCA Principal component analysis technique for feature extraction. PCA is very efficient for feature extraction because it reduces the dimension of vectors while ensuring essential information is not lost.

2.2 Colour based CBIR

The most commonly used global feature extraction techniques based on color are the MPEG-7 feature set such as dominant color descriptor, color histograms,color layout descriptor and scalable color descriptor

[4] Proposed a new color texture descriptor known as local binary pattern for color images(LBPC). The descriptor that is proposed is a two category method to set the get threshold for color pixels that are near the local window. Local binary patterns of the hue(LBPH) that in the HSL colorspace are local binary patterns of the hue component are derived to increase the discriminative power of the LBPC operator that is proposed here.LBPC+LBPH is derived from the fusion of LBPC, LBPH. An efficient image retrieval method LBPC+LBPH+CH. Is then derived by combining hue components(CH) to the fusion of LBPC, LBPH.

[5] The method proposed here uses the concept of color difference histogram(CDH). It aims at stimulating the sensitivity of the human eye towards the edge and colour feature of an image. This method retrieves the colour difference between two nearby pixels in terms of its orientation of colour and edge. The CDH is calculated in HSV colour space because it is closer to our color space.The color histogram in HSV color space is also used as a feature.Entropy and correlation criteria are used to select features that are efficient.

2.3 Shape based CBIR

The shape of objects has a basic role among different parts of visual data . Shape is an exceptionally fascinating feature when utilizing solidarity search and recovery. The edge-based shape interpretation methods are implemented . One of the most used operators Contour Detection Image processing.

[6] A Bi-layer Content Based Image Retrieval (BiCBIR) system has been proposed in this paper which consists of two modules: the first module extracts the features of dataset images in terms of color, texture and shape. Second module consists of two layers: initially all images are compared with the query image for shape and texture feature space and indexes of M most similar images to the query image are retrieved. Next, M images retrieved from the previous

layer are matched with query images for shape and color feature space and F images similar to the query image are returned as an output.

[7] presents a technique for CBIR (Content Based Image Retrieval) where the regions are selected on the basis of their contribution to image contents. At region-level Texture and edge features are extracted whereas at image-level shape features are extracted. The region-level image is to be divided in non overlapping regions. For each region texture and edge features are calculated separately. For extracting the texture feature Curvelet transform is used by providing continuity to the curve as well as line in the feature extraction process. Moment invariant is used for extracting the shape features. All of the regions may not have equal contribution in identifying the users perception of the image. Proposed techniques do not dominate the non-highlighted regions but it decreases the region weight for less contributing regions. IRM (Integrated Region Matching) method is used for retrieving the relevant images.

2.4 Composition based CBIR

Composition based CBIR uses hybrid methods for feature extraction.

[8] Due to recent advances in deep neural networks, new approaches to content-based image retrieval (CBIR) are developed. This proposed method presents a framework that learns a deep, dynamic metric between images using interactive CBIR. The proposed methodology is not only limited to hashes, precalculated categories, or clusters of the search space, but it is based on the user feedback. Here a deep learning framework that utilizes pre-extracted features from Convolutional Neural Networks is used and learns a new distance representation based on the user's relevance feedback.

[9] This paper proposes a new classifier named Extreme Learning Machine (ELM) on a hybrid framework for developing a Content Based Image Retrieval (CBIR) system to improve the accuracy problems faced with the earlier image retrieval system. The main aim of the system is to provide more accuracy and less consumption of time. This system uses Wang database with Local Binary Pattern (LBP), canny edge, color moment, and region props for the extraction of texture, color, edge and shape features respectively. For further implementation all the features from the image are extracted,to use it a distance matrix will be determined. To categorize all the

images, an ELM classifier is used in this proposed CBIR. For finding similar images Score Level Fusion, as similarity measure is used. The obtained results shows that the accuracy and efficiency of the CBIR system increased at a very rapidly after the use of ELM classifier in terms of f-measure, precision, recall, and retrieval time rather than just using similarity measure of the extraction features.

[10] Proposes a method based on new distance representation that uses convolutional Neural Networks and the relevance of learners.The CBIR system proposed here is an interactive system where the framework is able to understand naturally what features are suitable to address clients' issues.

[11] proposes a method to solve the problem of image retrieval. As the volume of data is large, they proposed the method using deep learning ,Deep Belief Network (DBN) to solve the problem. The method proposed is tested through correlation simulation and the results found showed huge positive deviation.

3. Overview of the project

3.1 Autoencoders

Autoencoders are a type of unsupervised learning algorithms and are trained on classless data. Autoencoders learn a compressed form of representation of input data by using neural networks. This project uses autoencoders to extract features of an image into a compressed form. LSR is a compressed form that contains the feature vector of an image. LSR of all the images in the database are computed and stored. When a query image is taken as input, its feature vector is calculated and compared with the feature vector of other images. Similarity between the feature vectors of query image and images in the database is calculated using k nearest neighbours classifier and LSR of k most similar images are found.Autoencoders consist of two parts: Encoder and decoder. Encoders convert all the input images into LSR while the decoder gets the original image back from the LSR representation.

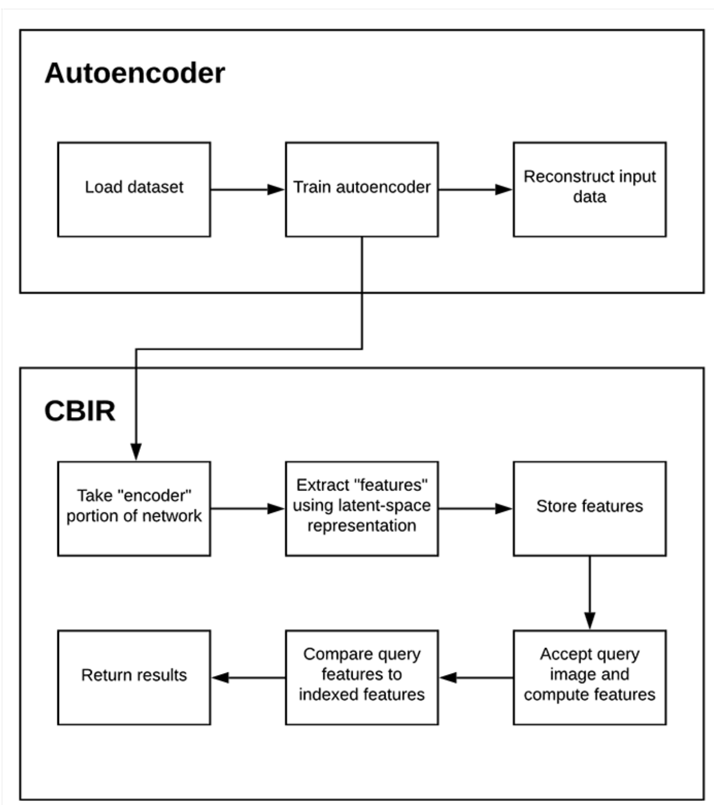


Fig.1 Content based image retrieval using autoencoders architecture
(Credit- https://ieeexplore.ieee.org/document/9138007)

3.2 Encoder

Encoders are used to produce LSR for every image. They extract features from an image and produce feature vectors as output. Features are extracted using CNN and stored as a feature vector in the final layer.

Encoders consists of following layers:

Layer 1 - 32 filters of size 3x3 and relu is used as activation function.
Layer 2 - Max Pool layer of size 2x2
Layer 3 - 64 filters of size 3x3 and relu is used as activation function
Layer 4 - Max Pool layer of size 2x2
Layer 5 - 128 filters of size 3x3 and relu is used as activation function
Layer 6 - Max Pool layer of size 2x2
Layer 7 - 256 filters of size 3x3 and relu is used as activation function
Layer 8 - Max Pool layer of size 2x2
Then output is flattened to convert the image to feature vector.
Layer 9 - Dense Layer

3.3 Decoder

After getting a LSR of similar images, these LSR are needed to be converted to the original image. Decoders take the LSR of an encoded image and generate the image back from LSR. It uses reverse methods to get the original image. It uses transpose convolution and elu activation function. It also consists of four layers.

Layer 1 - 128 filters of size kernel 3x3 and sigmoid is used as activation function.
Layer 2 - 64 filters of size kernel 3x3 and sigmoid is used as activation function
Layer 3 - 32 filters of size kernel 3x3 and sigmoid is used as activation function
Layer 4 - 3 filters of size kernel 3x3 and sigmoid is used as activation function

3.4 K nearest neighbors for similarity computation

When a query image is input, it is converted to LSR using encoder. Then similar images are searched from LSR of all images. Distance between feature vectors is used to get the similarity between images. Similar images are produced using nearest neighbours classifiers.k- Nearest neighbor classifier is used to get k similar images to the query image.

4. Implementation

4.1 Dataset

Dataset used is cifar 10 dataset. It consists of 50000 training images and 10000 testing images of size 32x32x3. Images consists of 10 classes.

4.2 Training

Before feeding the images to the model, the images are preprocessed, images are first divided by 255 to convert the value in range 0 to 1 and then subtracted by .5 to make it zero centered. Image shape height,width and color and code size i.e size of feature vector is given as input to model. Hyperparameters like number of epochs are initialized to 25, learning rate = 1e-3. The dataset is loaded and split into test and train data. The model is compiled using the adamax compiler and mean significant error(MSE) as a loss function. Once the model is trained we can use it to make predictions.

4.3 Testing

Model is tested using training and validation losses. Training and validation losses decrease with each epoch. The model becomes stable after 25 epochs. The final training and validation losses are .0114 and .0118 respectively. The given curve shows the model trains well and does not overfit.

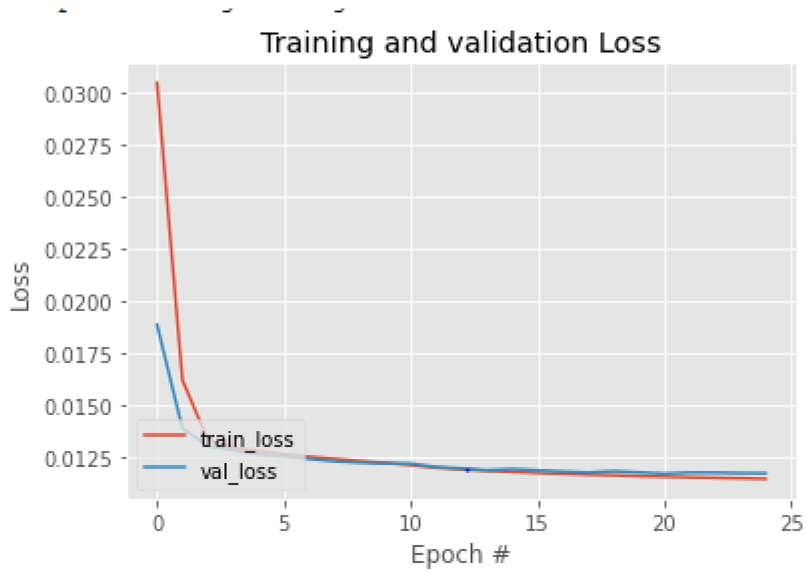
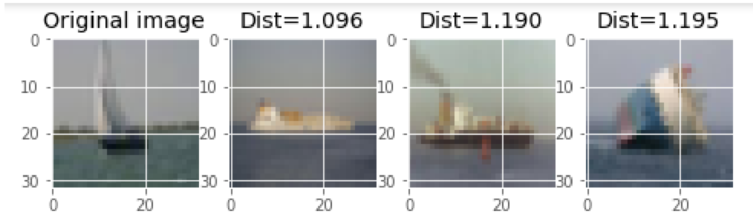
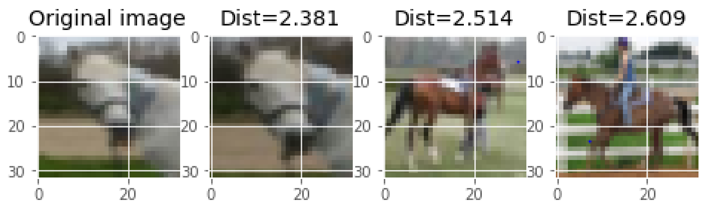


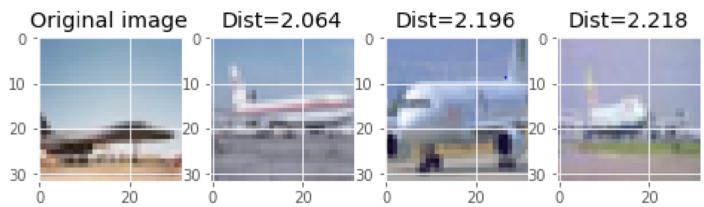
Fig.2 Training and Validation Loss curve

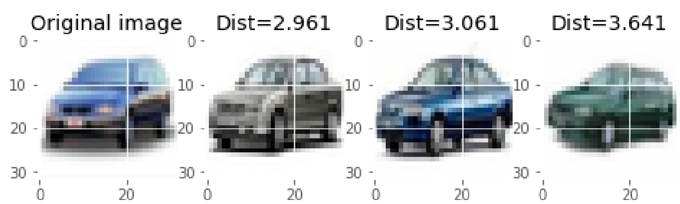
5. Results

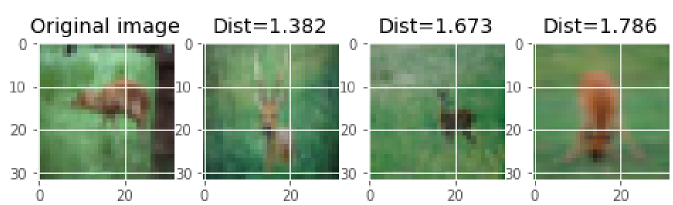
Results of the experiment. Original image shows the query image and 3 similar images are produced.











6. Conclusion

Autoencoder model for content based image retrieval is quite fast and efficient and achieves a good level of accuracy.The model uses unsupervised learning for producing feature vector and k nearest neighbor is used to get similar images based on similarity measures. The model is trained using a large dataset, after every epoch the training and validation loss decreases. After around 20 epochs it becomes steady. Although this method is quite fast, generating images back from LSR using deconvolution results in loss of information, thus the resultant image suffers from loss. Overall autoencoders based image retrieval produces quite good results and can be used in other search spaces.

REFERENCE

[1] Ch. Kodanda Ramu, T. SitaMahalakshmi‘An Efficient Image Retrieval System using GLCM Features and Kullback-Leibler Divergence’ International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 –8958, Volume-9 Issue-3, February 2020

[2] Sidi Ahmed M., Mohammed Amin Be., El Wardani D., Saïd Mahmoudi and Mohammed Be. , “Cloud-Based Image Retrieval Using GPU Platforms”, Computers 2019, 8, 48; doi:10.3390/computers8020048,2019.

[3] Shaimaa H. Shaker,Nuha M. Khassaf ‘Methods of Image Retrieval Based Cloud’, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-3, January 2020

[4] Singh, Chandan & Walia, Ekta & Kaur, Kanwal. (2018). Color Texture Description with Novel Local Binary Patterns for Effective Image Retrieval. Pattern Recognition. Volume-76.

[5] Qazanfari, Hamed & Hassanpour, Hamid & Qazanfari, Kazem. (2019). Content-Based Image Retrieval Using HSV Color Space Features. 13(10).

[6] Memon, Imran & Ali, Qasim & Pirzada, Nasrullah. (2017). A Novel Technique for Region-Based Features Similarity for Content-Based Image Retrieval. Mehran University Research Journal of Engineering & Technology. 37(2),383-396, 10.22581/muet1982.1802.14.

[7] Singh, S., Batra, S. An efficient bi-layer content based image retrieval system.Multimed Tools Appl 79, 17731–17759 (2020). https://doi.org/10.1007/s11042-019-08401-7

[8] Reddy, K. & R, Anandan & K, Kalaivani & Swaminathan, P.. (2018). A comprehensive survey on Content Based Image Retrieval system and its application in the medical domain. International Journal of Engineering and Technology (UAE). 7. 181-185.

[9] Dahiya, Kajol & Gautam, Gaurav. (2019). Content Based Image Retrieval System. International Journal of Computer Sciences and Engineering. 7. 470-475. 10.26438/ijcse/v7i6.470475.

[10] Pyykkö J., Głowacka D. (2017) Interactive Content-Based Image Retrieval with Deep Neural Networks. In: Gamberini L., Spagnolli A., Jacucci G., Blankertz B., Freeman J. (eds) Symbiotic Interaction. Symbiotic 2016. Lecture Notes in Computer Science, vol 9961. Springer, Cham

[11] Saritha, R.R., Paul, V. & Kumar, P.G. Content based image retrieval using deep learning process.Cluster Comput 22,4187–4200 (2019).

[12] Vaibhav Rupapara, Manideep Narra, Naresh Kumar Gonda, Kaushika Thipparthy & Swapnil Gandhi Auto-Encoders for Content-based Image Retrieval with its Implementation Using Handwritten Dataset, Proceedings of the Fifth International Conference on Communication and Electronics Systems (ICCES 2020) IEEE Conference Record # 48766; IEEE Xplore ISBN: 978-1-7281-5371-1