A Deep Learning Approach for Content-based Image Retrieval using Sparse Auto-Encoder

K Bharathi[1*], Dr. M. Chandra Mohan[2]

[1]*Associate Professor, Department of MCA, Loyola Academy, Secunderabad, India*
[2]*Professor of CSE, Jawaharlal Nehru Technology University Hyderabad, India*
[1]*Research Scholar, JNTU, Hyderabad, India*
Corresponding E-mail: kalva.bharathi@gmail.com

Abstract

Most search engines retrieve photographs using classic text-based algorithms that depend on descriptions and metadata. Content-based image retrieval (CBIR), image categorization, and investigation have all gotten a lot of attention in the previous two decades. High-level picture views are represented as feature vectors consisting of numerical values in CBIR and image classification-based algorithms.To isolate the main object from a picture, we first use segmentation and main object detection. The autoencoder is then used to extract features from the object and choose relevant features.Various deep learning representations are trained and tested, and the outcomes are compared to see which architecture maximises prediction scores while reducing computing costs in flower categorization identification. Images of flowers will be used to train and test models, with a selection of them being utilised for validation. Finally, the results of the experiments suggest that our technique can be used to search for images in a genuine picture database.

Keywords: Image Retrieval, Search Engine, Image Classification, Query Based Image, Deep Learning

1. Introduction

We frequently utilise search engines. We may use a search engine like Google to get the most relevant answers to our questions. The majority of the queries are text-based. However, most of the time, the text is really helpful in locating pertinent answers. For instance, suppose you want to look for a product on the internet, such as a t-shirt, but you don't know what it's called. How did you manage to track them down? So, you're free to write the shirt's description.The issue with using descriptions is that you'll end up with a large range of items. Worse, they will not be identical to the product you are looking for, necessitating the use of a more effective retrieval method.

To solve it, we may take the image of the product, extract its properties, and then utilise those features to find related items. This is referred to as content-based image retrieval. CBIR (content-based image retrieval) is a method of extracting relevant pictures from a single image. An image query and an image database make up the system.Using an action methodology, the system will begin by extracting features from all photos, whether the query or the image database. The algorithm will next compare the query to all photographs in the database to see how similar they are. In the end, the system will retrieve all the images that are greatly similar to the query. Numerous classification techniques, including linear prediction [16], autoregression (AR) models [17, 18], and spectral estimation [19], are common methods.

2. Literature Review

This section examines the most important extraction techniques employed in a study into the Content-based image retrieval systems. Duanmu [1] described Image retrieval using a method of Color moment invariant on the dataset COIL-100 with an accuracy of 0.985. Wang et al. [2] performed experiments on Content-based image retrieval with a method that Integrated colour and texture features for the dataset Corel obtained the results of accuracy of about 0.613. Zhang et al. [3] used a method of pseudo-Zernike invariants for the application of Object recognition on the dataset COIL-100. With an accuracy of 0.797, Guo et al. [4] tested an application on CBIR using a method error diffusion block truncation coding characteristics implemented on the database Corel. Shao et al. [5] worked on Image retrieval for MPEG-7 dominant colour descriptor for the data Corel and obtained the results level of 0.8964. Liu et al. [6] experimented with Region-based image retrieval with a method of High-level semantics using decision tree learning applied on the data Corel and achieved an accuracy of 0.768. Islam et al. [7] calculated an accuracy of 0.9767 for an application of Automatic categorization of image regions by a model Dominant colour-based vector quantization on the data Corel. With an accuracy of 0.97, Jiexian et al. [8] used a Multiscale distance coherence vector technique to perform Content-based picture retrieval on the MPEG-7 image database.

Papakostas et al. [9] used wavelet transformations and conducted experiments on various datasets. The classification capabilities of the wavelet moments are greatly improved when only effective characteristics are used in the feature selection approach. Because there are no specific datasets for content-based image retrieval, Liu et al. [10] conduct their experiments on Corel datasets.With 15000 photos, Corel-5000 and Corel-10000 are used to test retrieval performance in HSV, RGB, and Lab colour spaces. The study [11] mentioned texture based image retrieval systems about 10k colour images collected for different situations like people, garden, texture and landscapes etc. The suggested model oncolourand grey co-occurrence matrices approaches were compared for accuracy. The average precision and recall rates achieved by the embedded neural network with bandlet transform on the top 20 retrievals are compared to the other conventional CBIR [12] and the findings clearly beat other models. Irtaza and Jaffar [13] used the Corel datasets to demonstrate the usefulness of the SVM-based architecture and found that the findings were consistent.Fadaei et al. [14] conducted content-based picture retrieval studies on the Brodatz and Vistex datasets, which each contained 112 grayscale and 54 colour images. The distance between the query image and the dataset image is calculated, and the precision and recall rates are calculated using the photos with the least distance. Afshan Latif et al. [15] went over the methodologies of CBIR based on feature extraction in great depth. In their study publications, the writers also mentioned a major research gap among scholars.

3. Proposed System

3.1.Sparse Auto-encoders

An unsupervised model might be considered the deep learning-based CBIR with image retrieval. While training with an autoencoder, we didn't employ class labels. For each picture in our dataset, the autoencoder is subsequently utilised to construct the latent-space vector representation. The distance between latent-space vectors should then be calculated.

#1: Train the autoencoder

#2: Using the autoencoder, extract features from all of the photos in our collection by calculating their latent-space representations.

#3: To locate all related photos in the collection, compare latent-space vectors.

After successfully training our autoencoder(#1), we can continue on to the feature extraction/indexing component of the image retrieval process (#2). This stage requires us to utilise our trained autoencoder to accept an input image, do a forward pass, and then use the encoder component of the network's output to generate our index of feature vectors. The goal of these feature vectors is to measure each image's content.

4. Results and Discussion

Keras, a highly modular neural networks framework written in Python and capable of operating on top of either TensorFlow or Theano, is used to build the experiment.It is used as a personal computer with the configuration of Intel® Core™ i5-4310M CPU @ 2.70GHz × 4, 8 Gb RAM and 64-bit.The flowers-images data corpus was downloaded from the TensorFlow data repository with a size of 218MB of 3670 images. These images are categorized into 5 classes daisy, dandelion, roses, sunflowers, and tulips. All the images are preprocessed, loaded with a batch size of 32 and considered as 80% data for training and 20% for validation. Each of the three convolution blocks (tf.keras.layers.Conv2D) in the Sequential model has a max-pooling layer (tf.keras.layers.MaxPooling2D). A ReLU activation function ('relu') activates a fully linked layer (tf.keras.layers.Dense) with 128 units on top of it. There has been no tuning done to this model. Here, choose the Adam optimizer and Sparse Categorical Cross entropy loss function.

Figure1: Classes of Images – I

Figure2: Classes of images – II

The above figures1 and 2 are the sample pictures of the dataset. Figure3 shows the image Tulip extracted from the dataset.

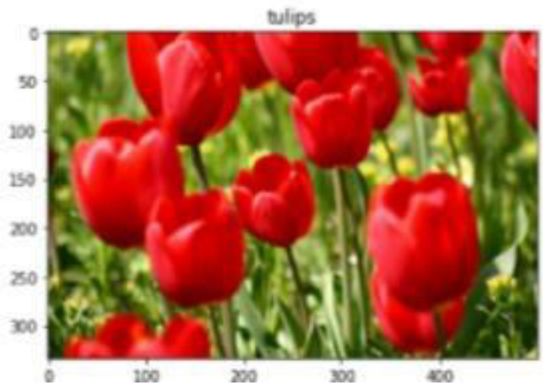

Figure3: Original image 'Tulip' extracted

After rescaling the above, standardized the values to the range of [0, 1] instead of [-1, 1]. The data augmentation is applied to the training data set for sub classifying, contrast, random, crop and colour inversion parameters. We are fitting the model by training the dataset, the accuracy is mentioned in the following figure 4.

```
Fit the model on training data
Epoch 1/5
92/92 [==============================] - 18s 171ms/step - loss: 1.3999 - accuracy: 0.3842 - val_loss: 1.1138 - val_accuracy: 0.5940
Epoch 2/5
92/92 [==============================] - 12s 121ms/step - loss: 1.0954 - accuracy: 0.5487 - val_loss: 1.0464 - val_accuracy: 0.5913
Epoch 3/5
92/92 [==============================] - 12s 121ms/step - loss: 0.9872 - accuracy: 0.6073 - val_loss: 0.9624 - val_accuracy: 0.6649
Epoch 4/5
92/92 [==============================] - 12s 126ms/step - loss: 0.9252 - accuracy: 0.6356 - val_loss: 0.9200 - val_accuracy: 0.6512
Epoch 5/5
92/92 [==============================] - 12s 124ms/step - loss: 0.8665 - accuracy: 0.6577 - val_loss: 0.8419 - val_accuracy: 0.6757
<keras.callbacks.History at 0x7fb37b533f10>
```

Figure4: The accuracy of training data

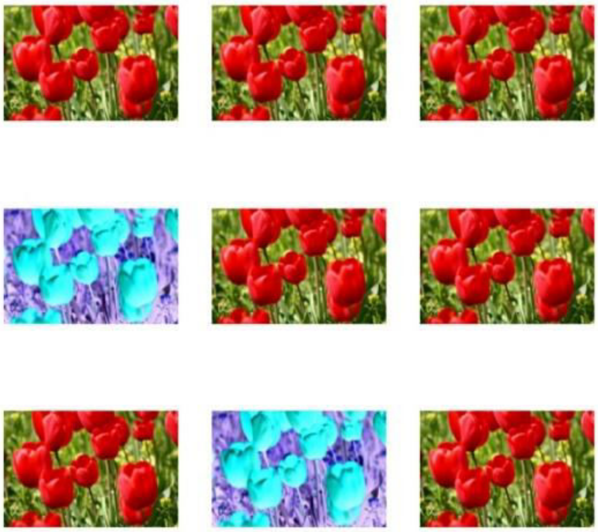

Figure5: Randomly inverted colour image.

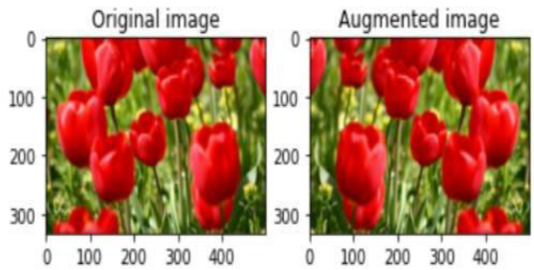

Figure6: Original image with augmented image

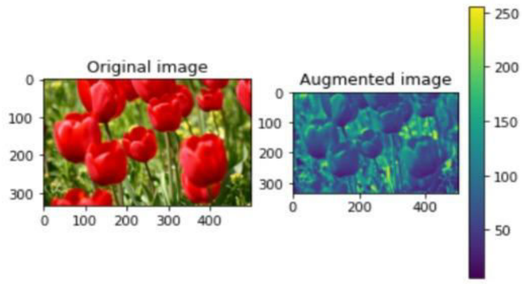

Figure7: Original image with grey scaled image

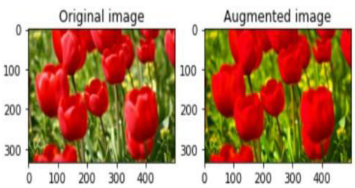

Figure8: Original image with saturated image

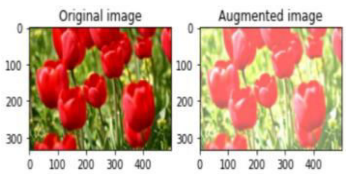

Figure9: Original image with brightness image

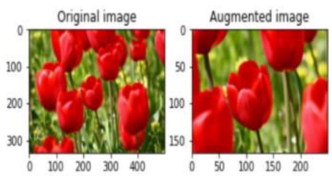

Figure10:Original image with cropped image

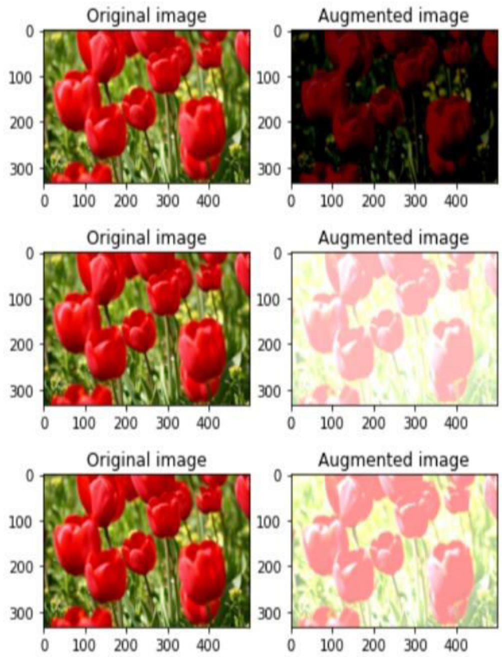

Figure11: Original image with randomly changed brightness

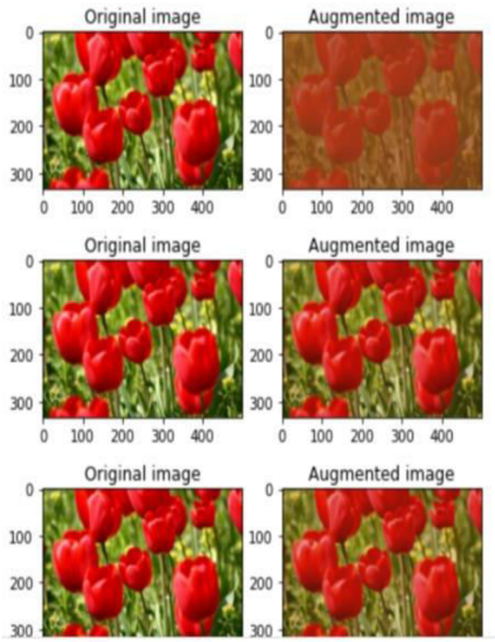

Figure12: Original image with randomly changed contrasts

5. Conclusion

We present a sparse encoder technique for a content-based picture retrieval system in this work. On a dataset of flowers, we tried a deep learning technique for picture retrieval. When it comes to the first recovered photos, our sparse autoencoder performs well. However, we put it to the test on photographs that were comparable. We also put the photographs through their paces in terms of colour, size, and rotation. We could obtain 94.37 percent accuracy by applying this.

References

[1] X. Duanmu, "Image retrieval using color moment invariant," in Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations (ITNG), pp. 200–203, IEEE, Las Vegas, NV, USA, April 2010.

[2] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," Multimedia Tools and Applications, vol. 68, no. 3, pp. 545–569, 2014.

[3] H. Zhang, Z. Dong, and H. Shu, "Object recognition by a complete set of pseudo-Zernike moment invariants," in Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), pp. 930–933, IEEE, Dallas, TX, USA, March 2010.

[4] J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 3, pp. 466–481, 2015.

[5] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning," Pattern Recognition, vol. 41, no. 8, pp. 2554–2570, 2008.

[6] M. M. Islam, D. Zhang, and G. Lu, "Automatic categorization of image regions using dominant color based vector quantization," in Proceedings of the Digital Image Computing: Techniques

and Applications, pp. 191–198, IEEE, Canberra, Australia, December 2008.

[7] Z. Jiexian, L. Xiupeng, and F. Yu, "Multiscale distance coherence vector algorithm for content-based image retrieval," @e Scientific World Journal, vol. 2014, Article ID 615973, 13 pages, 2014.

[8] G. Papakostas, D. Koulouriotis, and V. Tourassis, "Feature extraction based on wavelet moments and moment invariants in machine vision systems," in Human-Centric Machine Vision, InTech, London, UK, 2012.

[9] G.-H. Liu, Z.-Y. Li, L. Zhang, and Y. Xu, "Image retrieval based on micro-structure descriptor," Pattern Recognition, vol. 44, no. 9, pp. 2123–2133, 2011.

[10] X.-Y. Wang, Z.-F. Chen, and J.-J. Yun, "An effective method for color image retrieval based on texture," Computer Standards & Interfaces, vol. 34, no. 1, pp. 31–35, 2012.

[11] R. Ashraf, K. Bashir, A. Irtaza, and M. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," Entropy, vol. 17, no. 6, pp. 3552–3580, 2015.

A. Irtaza and M. A. Jaffar, "Categorical image retrieval through genetically optimized support vector machines (GOSVM) and hybrid texture features," Signal, Image and Video Processing, vol. 9, no. 7, pp. 1503–1519, 2015.

B. C. Chang and C. J. Lin, "LIBSVM: a library for support vector machines," ACM Transactions on Intelligent Systems and Technology (TIST), vol. 2, no. 3, pp. 1–27, 2011.

[12] S. Fadaei, R. Amirfattahi, and M. R. Ahmadzadeh, "Local derivative radial patterns: a new texture descriptor for content-based image retrieval," Signal Processing, vol. 137, pp. 274–286, 2017.

[13] Afshan Latif et al, Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review, Mathematical Problems in Engineering Volume 2019, 21 pages https://doi.org/10.1155/2019/9658350.

[14] V. Joshi, R.B. Pachori, A. Vijesh, Classification of ictal and seizure-free EEG signals using fractional linear prediction, Biomed. Signal Process. Control 9(2014) 1–5.

[15] E.D. Übeyli, Least squares support vector machine employing model-based methods coefficients for the analysis of EEG signals, Expert Syst. Appl. 37 (1)(2010) 233–239.

[16] S.-H. Kim, C. Faloutsos, H.-J. Yang, Coercively adjusted autoregression model for forecasting in epilepsy EEG, Comput. Math. Methods Med. 2013 (2013).

[17] K. Polat, S. Günes¸, Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals, Expert Syst. Appl. 34 (3) (2008) 2039–2048.