International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-11 Issue-9, August 2022

Deep CNN Based Hybrid Model for Image Retrieval

Amit Sharma, V.K. Singh, Pushpendra Singh

Abstract: ***The popularity of deep features based image retrieval and classification task has grown a lot in the recent years. Feature representation based on Convolutional Neural Networks (CNNs) found to be very effective in terms of accuracy by various researchers in the field of visual content based image retrieval. The features which are neutral to their domain knowledge with automatic learning capability from their images are in demand in various image applications. For improving accuracy and expressive power, pre-trained CNN models with the use of transfer learning can be utilized by training them on huge volume of datasets. In this paper, a hybrid model for image retrieval is being proposed by using pre-trained values of hyper parameters as input learning parameters. The performance of the model is being compared with existing pre-trained models showing higher performance on precision and recall parameters.***

Keywords: Content based Image Retrieval, Deep Convolutional Neural Network, Transfer Learning, Pre-trained Models.

I. INTRODUCTION

Starting through 1970's, the field of image retrieval gone through various phases of its advancements as initially it was based on text based search methods then content based search methods. There are lot of disadvantages of text based searches as they need high computational costs and lot of human efforts for finding relevant image information about the given query image. To overcome the limitations of text based search, the term 'content based image retrieval (CBIR)' was coined in 1992 by Japanese Electro technical engineer Toshikazu Kato [Wikipedia].CBIR consists of two steps such as feature extraction and similarity measurement. Feature extraction is based on low level features as colour, shape and texture of the image. Now-a-days, image retrieval methods uses low, high and deep features extracted from various deep learning methods [2]. The problem of combining low, high and deep features for feature extraction within image retrieval framework still remains open research task [2]. Traditional CBIR systems need handcrafted low level features such as colour histograms, SIFT, Tamura and many more. All handcrafted features faces problem of semantic gap with images as human perception always based on high level features such as objects, events etc.

Manuscript received on 09 July 2022 | Revised Manuscript received on 18 July 2022 | Manuscript Accepted on 15 August 2022 | Manuscript published on 30 August 2022.
***** Correspondence Author
Amit Sharma*, Research Scholar, Motherhood University, Roorkee (Uttarakhand), India.
Dr. V.K. Singh, Professor, Motherhood University, Roorkee (Uttarakhand), India.
Dr. Pushpendra Singh, Raj Kumar Goel Institute of Technology, Ghaziabad (Uttar Pradesh), India.

With the birth of Convolutional Neural Networks (CNN) in 1980s by Yann LeCun and its significant use in the field of image processing, feature extraction through CNN gains much popularity due to its superior performance in different tasks such as classification and recognition [3]. The reason of growth of CNN based feature extraction methods is availability of huge amount of image datasets already classified using pre-trained networks with GPU enabled modern computing machines [4]. The use of pre-trained models such as Alex Net, Google Net, Squeeze Net and ResNet-18 in computer vision, medical image classification and pattern recognition problems is very much popular these days with good performance. The concept of transfer learning with pre-trained models can be utilised for image retrieval problems requiring high volume of data generated daily and available for use.

In this study, a deep CNN based hybrid model is being proposed for image retrieval task and the performance of the model is being compared with existing pre-trained models showing higher performance on precision and recall parameters. The remainder of the paper is organised as follows: Section 2 refers to current state of art on CBIR research, section 3 refers to proposed methodology while experiment and results are reported in the section 4.Section 5 summarizes the findings and presents future directions of the paper.

II. LITERATURE REVIEW

The authors [1] proposed CIRPLANT model for composed image retrieval using nearest-neighbour approach on modified visual features generated through natural language processing. They use CIRR dataset of crowd source open domain images. The proposed model was compared on CIRR and existing fashion data Set and outperforms on recall parameter of image retrieval.

In the year 2021 paper [2] proposed a new image retrieval method as Deep Seated features histogram (DSFH), For extracting deep features, ranking whitening method was also proposed which reduces dimensions on FCT layer of VGG-16 pre-trained model to achieve higher precision of retrieval images. The authors used Oxford 5k and Paris6k datasets for their proposed models and compare precision and recall values on Corel10k and GHIM10k datasets also. Their proposed method was able to discriminate low level features and can match similar styles of scenes in the CBIR framework. The authors suggested to use any hybrid model for image retrieval which can combine low and deep features during feature extraction phase in the near future for generalizing their proposed model.

The paper [3] proposed a hybrid model using auto encoder network for dimension reduction and deep CNN to extract high level semantic features. The Similarity between dataset and query image was calculated using Annoy algorithm on CIFAR-10 and MNIST datasets. The author designed a 6 layer convolutional, 2- layer full-convolution CNN to extract deep features and 3-layer auto-encoder for dimension reduction getting a 128-bit vector for feature representation. The query engine was generated by Annoy algorithm and a tree structure was used to search and store index values of images. The hybrid model achieves almost 100% accuracy for MNIST dataset but lacks accuracy for CIFAR10 dataset. The authors suggest to use any hash algorithm for similarity measurement to improve accuracy in future on any large image dataset. In [4], a CBIR system based on pre-trained models was developed using transfer learning concept for extracting automated deep features independent of domain knowledge having low semantic gap with respect to human visual perception. The author considered Corel-1x and GHIM-10 k datasets for ResNet18 features achieving overall 95.5% and 93.9 % accuracy respectively which was superior to state of art CBIR systems using hand crafted features for feature extraction. The authors suggested using more combinations of pre-trained models to achieve high accuracy in the future on real time datasets also.

The paper [5] reviewed deep learning based CBIR on parameters as different supervision, different network, different descriptor and different retrieval type. This review shows superior performance of image retrieval with the use of generative adversarial networks, auto encoder networks and reinforcement learning networks or by using combination of these networks. The paper also shows the use of optimized objective functions as a trend in improved retrieval framework. The authors suggested using better deep models, specific objective functions, semantic preserving feature learning and attention based features for learning. Any retrieval system can be distinguished based on discriminative ability, fast image search and robustness capability. Self-supervised learning can be paradigm for optimizing retrieval performance in future.

An Adadelta optimized residual network for parameter tuning based CBIR model was proposed involving ResNet50 based feature extraction in [6]. A self-supervised model having multiple query images and attention based feature extraction architecture was proposed and shown improved retrieval performance on road scene datasets [7]. The authors suggested to use local aspects of images for further improvement in the performances of image retrieval system [7].

III. RESEARCH METHODOLOGY

To perform image retrieval based on deep CNN models, the proposed method has been divided in to five components such as block diagram of proposed deep CNN based hybrid model, architecture of deep CNN model, automatic deep feature extraction using pre-trained models, performance evaluation and validation, and dataset Description.

A. Block Diagram of Proposed Deep CNN Based Hybrid Model

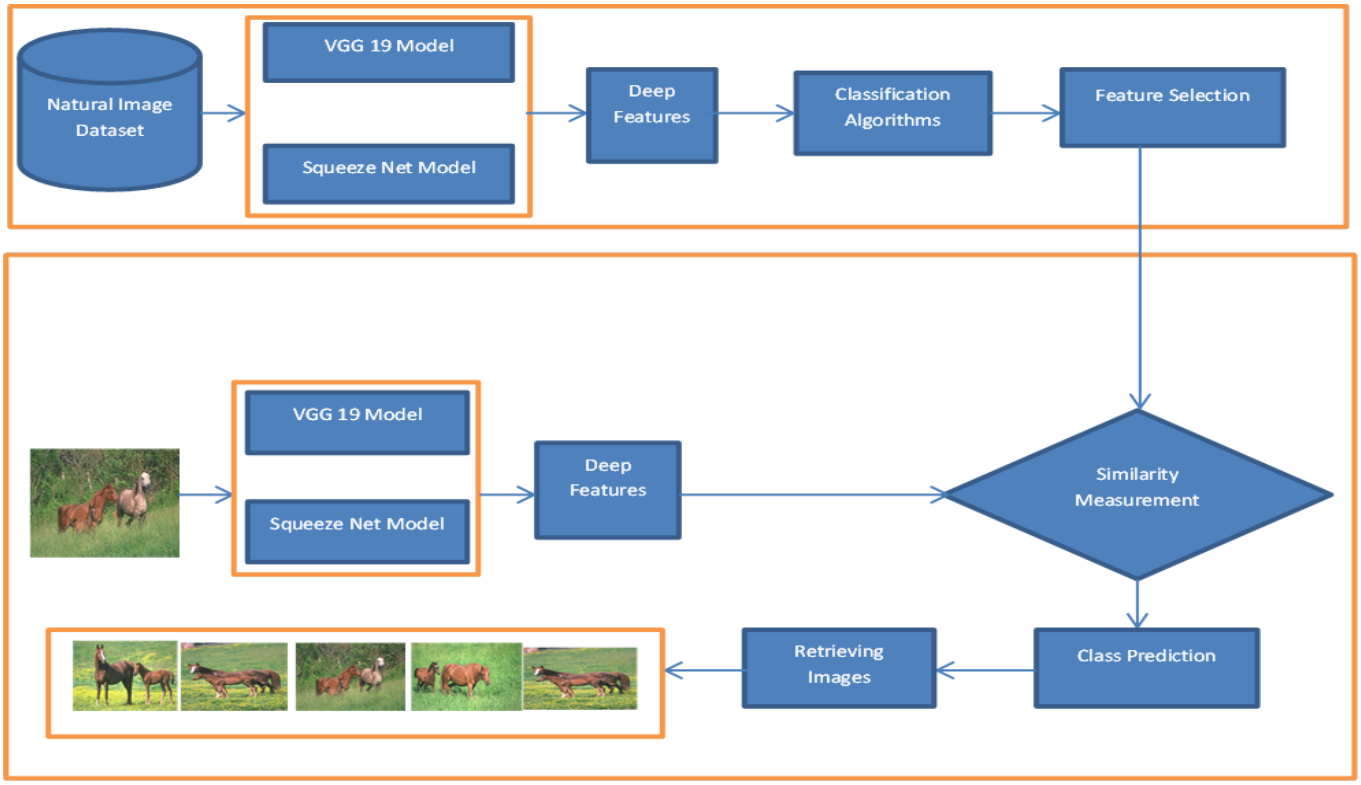

Figure1. Research Methodology

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The structure of neural cells in the human visual brain looks like linear bar or edge shaped receptive fields implying the possibility to use neural network models for low and deep feature extraction for images. Figure 1 demonstrates the block diagram of proposed model for visual content based image retrieval.The first component defines and extracts visual features using deep CNN based algorithms for selected natural image dataset. The feature set is used for feature selection with the help of classification algorithms at fully connected layer of deep CNN model. This offline process of deep feature extraction and feature selection results a set of deep feature which will be used during similarity matching with query image in place of database images. The same process is adopted for generation of query image deep features for similarity matching as input features. The standard Euclidian distance similarity metric is used for calculating similarity between query and dataset images. The output of the retrieval process is presented in the order of high similarity normalized value as 0 and lowest similarity normalized values as 1. Top 10 most similar images are displayed as a result of hybrid model of image retrieval.

B. Deep Convolution Neural Network Model

The proposed CNN model takes images as input to the first convolutional layer in the form of height, width and depth where depth of the image can be defined as 3rd dimension of network layer's activation volume. The neurons of any layer can be connected to a small region of the previous layer in the network. A CNN model can be defined as a sequence of many layers which can pass one volume of image data to another using various differentiable functions. The architecture of CNN model can stack convolutional, pooling, activation (ReLu, Softmax) and fully connected layers on top of each other for being operational for different tasks. For classification problems, a softmax activation layer is used to get the probabilities of different classes available in the dataset.

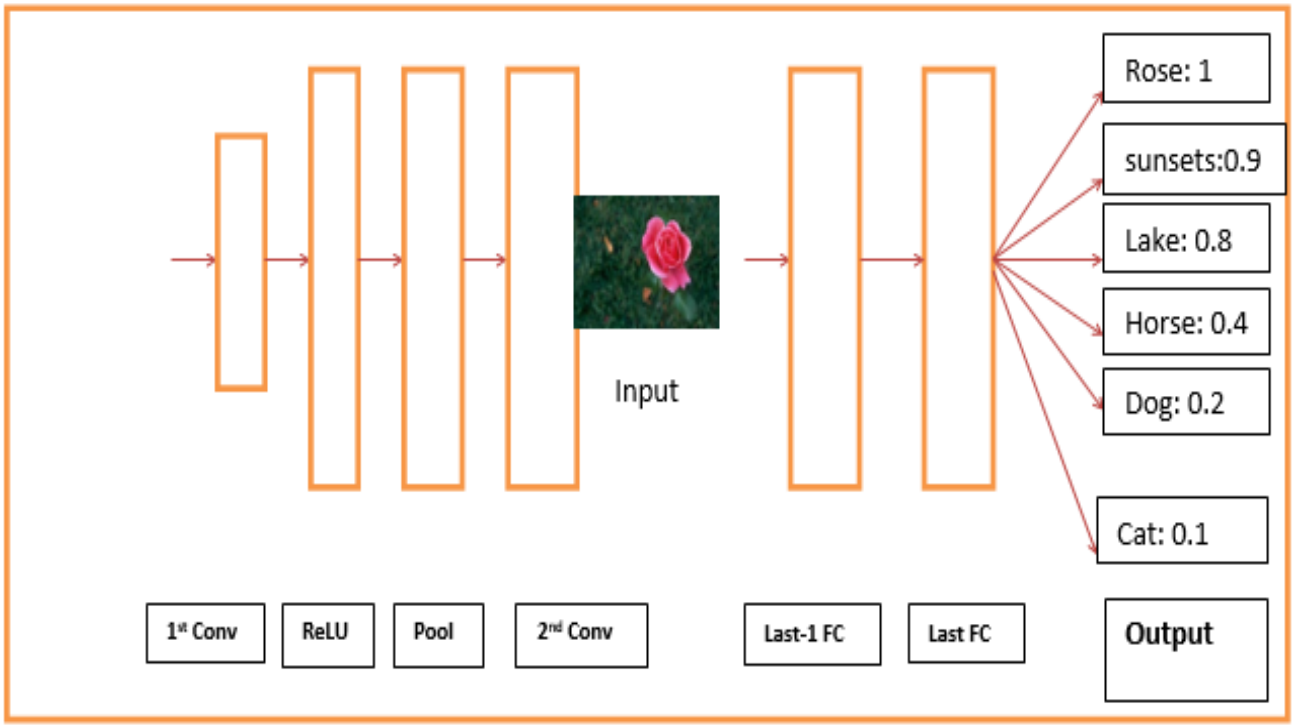

Figure 2. Proposed Deep CNN Model

C. Automatic Feature Extraction using Pre-trained Networks

a. Visual Geometry Group (VGG) – 19 Pre-trained Network

This pre-trained network originally builds and trained by Karen Simonyan and Andrew Zisserman at the University of Oxford in 2014 consists of 19 layers with input image size 224x224x3 having accuracy around 90% with top 5 images as a result. This model can classify 1000 objects from 1000 different classes from the image net dataset.

b. SqueezeNet Pre-trained Network

It is a pre-trained CNN model trained on large volume of Image Net dataset with 18 layers. It can classify any dataset up to 1000 class labels and process the input image size as 227x227x3.When compared with AlexNet, the size of SqueezeNet is less than 50% with equivalent accuracy for top 5 images. After training, model pruning and quantization of parameters can be applied to the SqueezeNet to reduce parameter size from 5MB to 500KB. This pre-trained model is known as resource efficient for variety of computer vision applications.

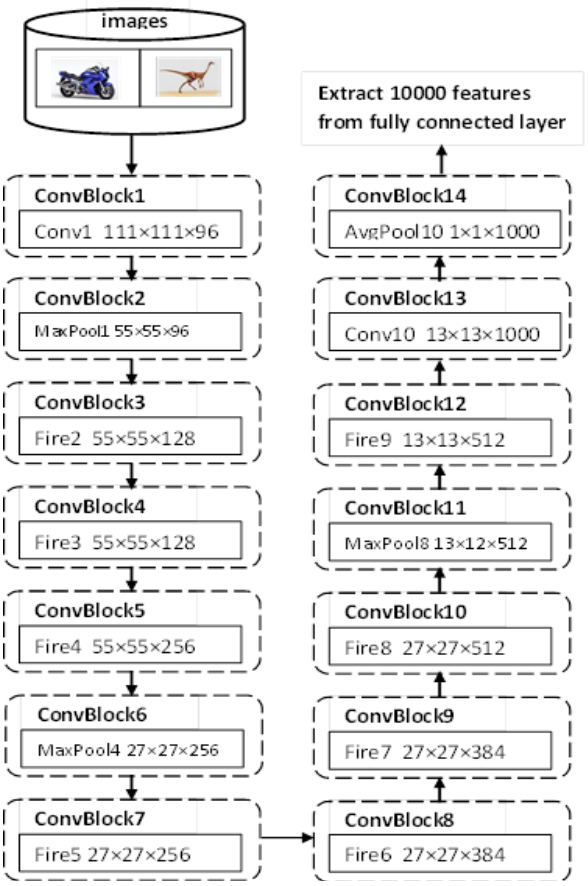

Figure 3. SqueezeNet Model [4]

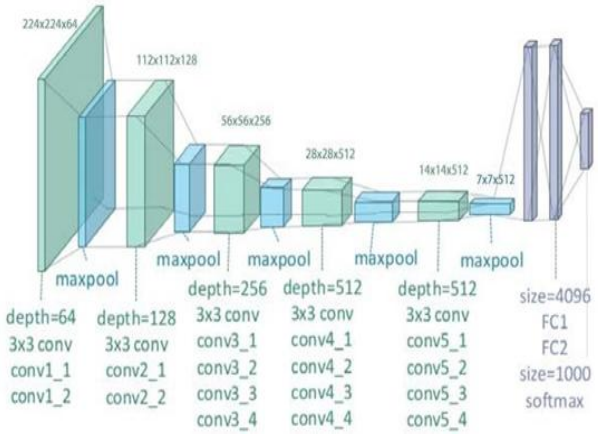

Figure 4. TheVGG-19 Network [15]

D. Performance Evaluation and Validation

The performance of any image retrieval system can be measured using various standard parameters. As per the literature there is no single parameter for evaluating all types of retrieval systems. Any evaluation parameter will depend on requirement of the user, selection of algorithm and domain of the problem. Based on these requirements, following parameters can be used to evaluate the performance of a deep CNN based image retrieval system.

1. Precision (p)

It can be defined as the ratio between the number of relevant images and total number of retrieved images.

$$p = \frac{N_p}{N_p + F_p}$$

Where Np = Number of relevant images Fp= Misclassified images as relevant (i.e. False positive)

2. Recall(r)

It is the ratio of relevant retrieved images to the total number of relevant images in the database.

$$r = \frac{N_p}{N_p + F_n}$$

Where F_n = False negative images i.e. the images actually belongs to relevant class but misclassified to any other class.

3. F-Measure

It can be defined as harmonic mean of precision and recall. The higher value of f-measure shows higher prediction capability of the retrieval system.

$$F = 2.\frac{p * r}{p + r}$$

E. Dataset Description

The proposed hybrid model for image retrieval is tested on a natural image dataset. The images of the dataset were collected from Google and from freely available datasets. All the collected images go through pre-processing phase before they are considered as input images. The whole dataset contains 20 distinct classes having 1000 images in each class. The dataset has been divided into training and test data in the ratio 80:20. The input image dimensions used by the pre-trained networks are 224x224x3 and 227x227x3.

IV. RESULTS AND DISCUSSION

The process of image retrieval has been implemented using Keras, Scikit-learn and OpenCV libraries in Python programming language for training and testing of deep CNN based proposed hybrid model. The retrieval process contains two steps: Deep feature extraction, classification and feature selection are carried out in the first step and similarity measurement using Euclidian distance in the second step. A deep feature vector containing concatenated features from two different pre-trained networks are used as input for feature classification and selection step. Proposed hybrid model uses weight and architecture of VGG19 and Squeeze Net for image classification. Both pre-trained models uses same dataset for training and testing in 80:20 ratio.

Retrieval Number: 100.1/ijitee.G92030811922
DOI: 10.35940/ijitee.G9203.0811922
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Figure 5. Samples Images from Natural Image Dataset [20 classes]

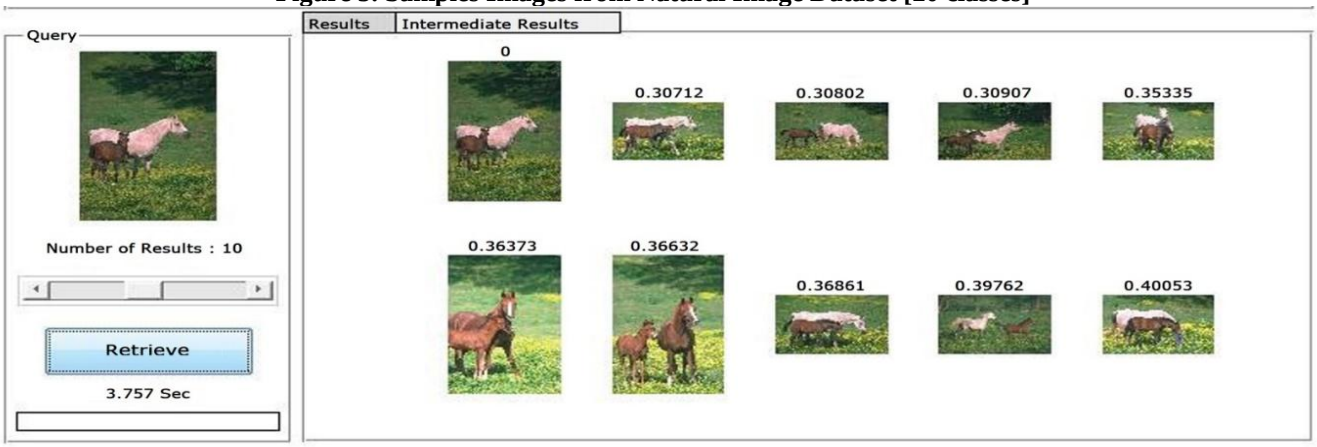

Figure 6. Top 10 retrieval results from Natural Scene Dataset

The precision and recall values are calculated using the formula's used in the definition of these terms. The sample image from each class is shown in Figure 5. The hybrid model retrieves top 10 most similar images with query image shown in the right panel of Figure 6 having value 0 as most similar image and 1 as most dissimilar images.

Table i. Average recall and precision for natural image dataset

Class	Squeeze Net Features		VGG 19 Features		Hybrid Features [Squeeze Net & VGG19]	
	Recall (R)	Precision (P)	Recall (R)	Precision (P)	Recall (R)	Precision (P)
Arabian_Horses	0.070	0.72	0.082	0.82	**0.085**	**0.86**
Aviation_Photography	0.066	0.67	0.085	0.85	0.087	**0.88**
Barns_and_Farms	0.080	0.81	0.090	0.90	0.092	**0.92**
Bears	0.081	0.82	0.091	0.91	0.090	**0.90**
Buses	0.085	0.85	0.092	0.92	0.093	**0.93**
car_old2	0.055	0.56	0.070	0.70	0.080	**0.80**
castles	0.061	0.62	0.077	0.77	0.082	**0.83**
Cats	0.088	0.89	0.093	0.93	0.095	**0.95**
Dogs	0.096	0.96	0.099	0.99	0.099	**0.99**
drinks	0.087	0.87	0.098	0.98	0.098	**0.98**
Flowers_2	0.091	0.91	0.095	0.95	0.096	**0.96**
Lakes_and_Rivers	0.075	0.75	0.085	0.85	0.087	**0.87**
Roses	0.063	0.63	0.080	0.80	0.081	**0.82**
Sailboats	0.076	0.76	0.090	0.90	0.094	**0.94**
Sports	0.080	0.80	0.085	0.85	0.088	**0.88**
Sunsets	0.071	0.71	0.080	0.80	0.082	**0.82**
Tigers	0.080	0.80	0.091	0.91	0.093	**0.93**
Trees_and_Leaves	0.069	0.70	0.082	0.82	0.083	**0.83**
Tulips	0.076	0.76	0.092	0.92	0.094	**0.94**
Waterfalls	0.087	0.87	0.098	0.98	0.099	**0.99**
Mean	**0.07685**	**0.773**	**0.08775**	**0.8775**	**0.0899**	**0.901**

V. CONCLUSION AND FUTURE WORK

In summary, a hybrid model for deep feature based visual content based image retrieval has been proposed in this paper. For feature extraction and classification, two pre-trained models namely VGG19 and Squueze Net have been considered for concatenated deep feature vector on a natural image dataset. The selected dataset images were collected from freely available search engines and pre-processed to convert into suitable resolutions as required by pre-trained networks for input images. For measuring similarity, Euclidian distance as parameter is used. The performance of retrieval system is evaluated based on precision and recall parameters. Based on the results from retrieval system, It could be said that hybridisation of pre-trained networks for feature extraction step could improve the values of precision and recall as shown in the above Table I. Improved values of these parameters can directly be inferred as better and accurate retrieved top 10 images. In future, hybridisation of more than two pre-trained models on different datasets of large volumes can be done for improving the performance of image retrieval system based on deep CNN models.

REFERENCES

1. Z. Liu, C. Rodriguez-Opazo, D. Teney and S. Gould, "Image Retrieval on Real-life Images with Pre-trained Vision-and-Language Models," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 2105-2114, doi: 10.1109/ICCV48922.2021.00213. [CrossRef]
2. Guang-Hai Liu, Jing-Yu Yang, "Deep-seated features histogram: A novel image retrieval method", *Pattern Recognition, Volume 116, 2021, 107926, ISSN 0031-3203,* https://doi.org/10.1016/j.patcog.2021.107926. [CrossRef]
3. J. Qin, E. Haihong, M. Song and Z. Ren, "Image Retrieval Based on a Hybrid Model of Deep Convolutional Encoder," *2018 IEEE International Conference of Intelligent Robotic and Control Engineering (IRCE)*, 2018, pp. 257-262, doi: 10.1109/IRCE.2018.8492952. [CrossRef]
4. Ali Ahmed, "Pre-trained CNNs Models for Content based Image Retrieval" International Journal of Advanced Computer Science and Applications(IJACSA), 12(7), 2021. http://dx.doi.org/10.14569/IJACSA.2021.0120723. [CrossRef]
5. S. R. Dubey, "A Decade Survey of Content Based Image Retrieval Using Deep Learning," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 5, pp. 2687-2704, May 2022, doi: 10.1109/TCSVT.2021.3080920. [CrossRef]
6. M. P, K. S, S. V and S. A, "Intelligent Content Based Image Retrieval Model Using Adadelta Optimized Residual Network," *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2021, pp. 1-5, doi: 10.1109/ICSCAN53069.2021.9526470. [CrossRef]
7. H. Govindarajan, P. Lindskog, D. Lundström, A. Olmin, J. Roll and F. Lindsten, "Self-Supervised Representation Learning for Content Based Image Retrieval of Complex Scenes," *2021 IEEE Intelligent Vehicles Symposium Workshops (IV Workshops)*, 2021, pp. 249-256, doi: 10.1109/IVWorkshops54471.2021.9669246. [CrossRef]
8. T. Sunitha, T.S. Sivarani, "Novel content based medical image retrieval based on BoVW classification method", *Biomedical Signal Processing and Control,Volume 77, 2022, 103678, ISSN 1746-8094,* https://doi.org/10.1016/j.bspc.2022.103678. [CrossRef]
9. Singh, Pushpendra and Hrisheekesha, P.N. and Singh, V.K., Comparative Study of Color and Texture Features for Image Retrieval on Natural Datasets (March 12, 2019). Proceedings of 2nd International Conference on Advanced Computing and Software Engineering (ICACSE) 2019, Available at SSRN: http://dx.doi.org/10.2139/ssrn.3350992. [CrossRef]
10. Jia Chen, Haidongqing Yuan,Yi Zhang, Ruhan Heand Jinxing Liang, "DCR-Net: Dilated convolutional residual network for fashion image retrieval", Computer Animation and Virtual Worlds,16 May 2022, https://doi.org/10.1002/cav.2050. [CrossRef]
11. Singh Pushpendra, Hrisheekesha P.N. and Singh Kumar Vinai, "CBIR-CNN: Content-Based Image Retrieval on Celebrity Data Using Deep Convolution Neural Network", *Recent Advances in Computer Science and Communications 2021; 14(1)* https://dx.doi.org/10.2174/2666255813666200129111928. [CrossRef]
12. Anbang Yao and Shan Yu, "Robust Face Representation using Hybrid Spatial Feature Independent Matrix, *IEEE Transactions on Image Processing, Vol.22, No.8, pp.3247-3259,* August, 2013. [CrossRef]
13. Singh Pushpendra, Hrisheekesha P.N. and Singh K. Vinai, "Ensemble Visual Content Based Search and Retrieval for Natural Scene Images", *Recent Advances in Computer Science and Communications 2021; 14(2).* https://dx.doi.org/10.2174/2213275912666190327175712. [CrossRef]
14. Rong-Xiang Hu, et.al, "Angular Pattern and Binary Angular Pattern for Shape Retrieval", *IEEE Transactions on Image Processing, Vol. 23, No. 3, pp. 1118-1127,* MARCH 2014. [CrossRef]
15. Pushpendra Singh, V K Gupta and P N Hrisheekesha. "A Review on Shape based Descriptors for Image Retrieval", *International Journal of Computer Applications* 125(10):27-32, September 2015. doi: 10.5120/ijca2015906043. [CrossRef]
16. Hao Xia, "Online Multiple Kernel Similarity Learning for Visual Search", *IEEE Transactions on Image Processing, Vol. 36, No. 3, pp. 536-549,* MARCH 2014. [CrossRef]

AUTHOR PROFILES

Amit Sharma, I am currently pursuing Ph.D. from Motherhood University, Roorkee, UK. I have done M.Tech.(CSE) and B.Tech (CSE). I have worked with various reputed engineering colleges as assistant professor in the department of computer science and engineering. I have taught many subjects like C Programming, Soft Computing, Machine Learning and Data Analytics at the under graduate level students. I have guided more than 30 M.Tech and B.Tech student's projects in the field of Soft Computing and Image Processing. I have authored 10 research papers in different international journals and conferences. My research interests include Image Processing, Deep Learning and Soft Computing. Email id: amit.faculty@gmail.com

Prof (Dr.) V. K. Singh, is currently working as Director Research at Motherhood University Roorkee, UK. He has completed his Ph.D. in Mathematics from Indian Institute of Technology, Varanasi, U.P (INDIA) in 2001. He has worked with many reputed engineering colleges on different roles as Director, Dean And Professor since more than 20 years. He has authored more than 30 research papers with SCIE and Scopus indexed journals, authored four books with different publishers such as Springer and worked as Convener in series of Springer international conferences on Modern Mathematical Methods and High Performance Computing in Science and Technology. He has also completed one DST project on Mathematical Modelling and Simulation of data using high performance computing in 2012-14.His research area includes Image Processing, Optimization, Approximation Theory and Functional Analysis in Computational Mathematics. (Email id: drvinaiksingh@gmail.com)

Dr. Pushpendra Singh, He is currently working as Associate Professor & Head, in the department of Information Technology at Raj Kumar Goel Institute of Technology, Ghaziabad, Uttar Pradesh (INDIA. He has completed Ph.D. (CSE) in the field of image processing from Dr. APJ Abdul Kalam Technical University, Lucknow (UP)in June 2021, Master of Engineering (CSE) from NITTTR, Chandigarh in 2012 and B.Tech (CSE) from Institute of Integral Technology (Now Integral University) Lucknow in 2004.He has authored over 20 (4 SCIE, 5 Scopus and 8 peer-reviewed) journal and conference articles and 4 patents (1 Australian patent Granted and 3 Indian Patents Published). His research interests include computer vision, pattern recognition and Deep Learning. Email id: pushpendra.singh1@gmail.com