

Image Retrieval System using Residual Neural Network in a Distributed Environment

R Rajkumar, M V Sudhamani

Abstract: Development of Content-Based Image Retrieval systems supports retrieval of similar images based on selected features. Selection of appropriate features for this process is a difficult task. In this regard, deep learning concept helps in choosing appropriate features for retrieval. In this work, Content-Based Image Retrieval system is proposed using Convolution Neural Network known as Residual Neural Network model. The dataset used to build retrieval system is collection of web images 50,000 of 250 categories. The model is trained on 40% of image data and tested on 60% of data. When user submits a query image from the client-side, similar features are extracted by the model on server-side. Later, the features of query image are compared with trained images data and similarity is measured using the metric of Euclidean distance. The retrieved resultant images are displayed on Graphical User Interface. The results are comparatively higher with the existing systems. The proposed work is also compared with Google’s Image retrieval system for random query images and our proposed work has shown a better performance by 14.27%.

Keywords: Content-Based Image Retrieval, Residual Network, Convolution Neural Network, Euclidean distance.

I. INTRODUCTION

In current scenario of growth rate in digital images on the world wide web, it is a complex process to search for images of similar type, which is one of the major active topics in computer vision area. With the traditional systems, retrieval of similar images is performed by selecting the local and / or global features of an image like SIFT [3] or BoVW [4]. These traditional image retrieval systems perform well for images when the number of images is less in number. But for large image repositories or collections, these methods do not show a higher accuracy of retrieval of similar images [1]. The recent methods like conventional neural network (CNNs) uses unsupervised learning techniques and the model is trained with several parameters like learning rate, error rate, loss function, number of epochs the model is to be trained, etc [2]. These models have shown a better performance in terms of classification of images and retrieval of similar images from large image collections or repositories.

The neural networks consist of three layers namely input layer, hidden layer and output layer as shown in figure 1. The hidden layer can have more than one layer for performing different operations like filtering, edge detections, color extraction, corners detection, texture recognition, etc., are applied on images to extract these features. Such hidden layers can be enhanced to a larger number in order to get higher performance results.

Such a model that has several layers to form a deep network is termed as Deep Convolution Neural Networks (DCNNs) or Deep learning models. These DCNNs can be used as feature extractors in order to match or compare the features of query and repository images during similar image retrievals.

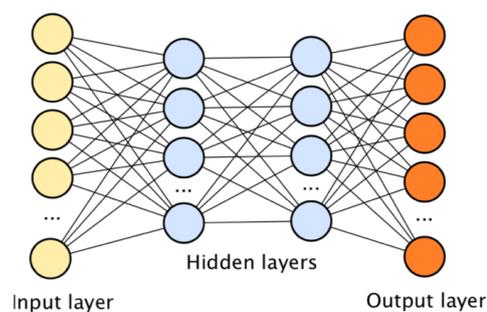


Figure 1 General Convolutional Neural Network

In deep networks, as the number of hidden layers is increased, then there can be a point beyond which if any layers are added the performance may not increase further and reaches a saturation or there can be a degradation in the performances too. Such a limitation is a bottleneck for further improving the network model by adding more layers. Such degradation of the system is due to vanishing gradients problem. This degradation can be overcome using skip-connections model called Residual Neural Network model (ResNet) and was proposed by Microsoft Research Team [5] and is the winner of ImageNet challenge during 2015. In this paper, we have proposed the work using the Residual Neural Network that uses deep learning model that have a total of 18 layers and hence referred to as ResNet18 model.

The rest of the paper is organized as: in section II the related works have been described in brief and section III presents the proposed architecture and technique in detail, while in section IV experimental results and discussions are given. Finally, in section V, conclusions are drawn and some ideas for future work is discussed.

II. RELATED WORKS

The content-based image retrieval (CBIR) system are implemented using different techniques in the recent decades.

Manuscript received on March 15, 2020.

Revised Manuscript received on March 24, 2020.

Manuscript published on March 30, 2020.

* Correspondence Author

Mr. R Rajkumar, Research Scholar-VTU, Asst. Professor, Information Science & Engg., RNS Institute of Technology, Bengaluru, India. raj@rnsit.ac.in

Dr.M.V.Sudhamani Professor & HoD, Department of Information Science & Engg., RNS Institute of Technology, Bengaluru, India. mvsudha_raj@hotmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The CBIR systems uses different features available in images like color, shape and textures [9, 10, 11, 12] and few of these features are invariant to local transformations or scaling [7]. The bag of visual words (BoVW) as described in [3] are used for retrieval of images with a histogram of color bins that form the feature bags or patterns for the whole dataset. Deep Residual Networks is used along with kernel-based supervised hashing to form a framework in [14]. This framework uses the learning capabilities of ResNet to mine deep features that are hidden in the image and extract them as feature descriptors, thereby enhancing the visual expressions of images. On the other hand, the kernel-based supervised hashing learns from the high-dimensional image feature and map into low-dimensional hamming space and resulting in compact Hash codes. These codes are used in image retrieval using low-dimensional hamming space. The datasets used to carry out the experiments are MNIST, CIFAR-10 & 100, Caltech 256.

Training of the CNN models requires a large collection of images as well information associated with the images in the form of annotations. These annotations are tedious task when any human needs to accomplish it by manually filling the description related to hundreds of thousands of images in a collection. Hence, there needs a mechanism to automate such a tedious work using CNN model or fine tuning of existing system [15]. The fine tuning of CNN is proposed using a reconstructed 3D models obtained using retrieval

and structure-from-motion methods. This helps in guiding the selection using training data. It is observed that in this work, descriptor whitening discriminatively learned by CNN model has outperformed normal PCA whitening. A novel trainable Generalized-Mean (GeM) pooling layer is proposed that generalizes the max-pool and average pool techniques and boosts retrieval performance on VGG network.

Image classification using different deep learning techniques is briefed with various datasets [16]. Comparison of the transfer learning, CNN deep learning and others are listed for experiments conducted for AlexNet, GoogleNet, ResNet, etc., providing an insight of deep learning techniques.

The multimodal retrieval technique is explored with single modality query and retrieval from a collection of multiple modalities. The technique uses two layers, embedding and relevance matching layers, that is used for training and matching of images. Further, nearest neighbour is used to retrieve relevant images [17].

The deep learning methods are applied on medical images too, in which the dimensions of the images 2D and 3D, the CNNs applied on such images are termed as 2D, 2.5D and 3D CNN models [18]. The CNNs with 2D images uses 2D filters and in case of 2.5D CNNs, it provides detailed spatial information of neighbouring pixels and reduces the training cost as compared with 3D model.

III. ARCHITECTURE AND TECHNIQUES

A. Architecture of the proposed work

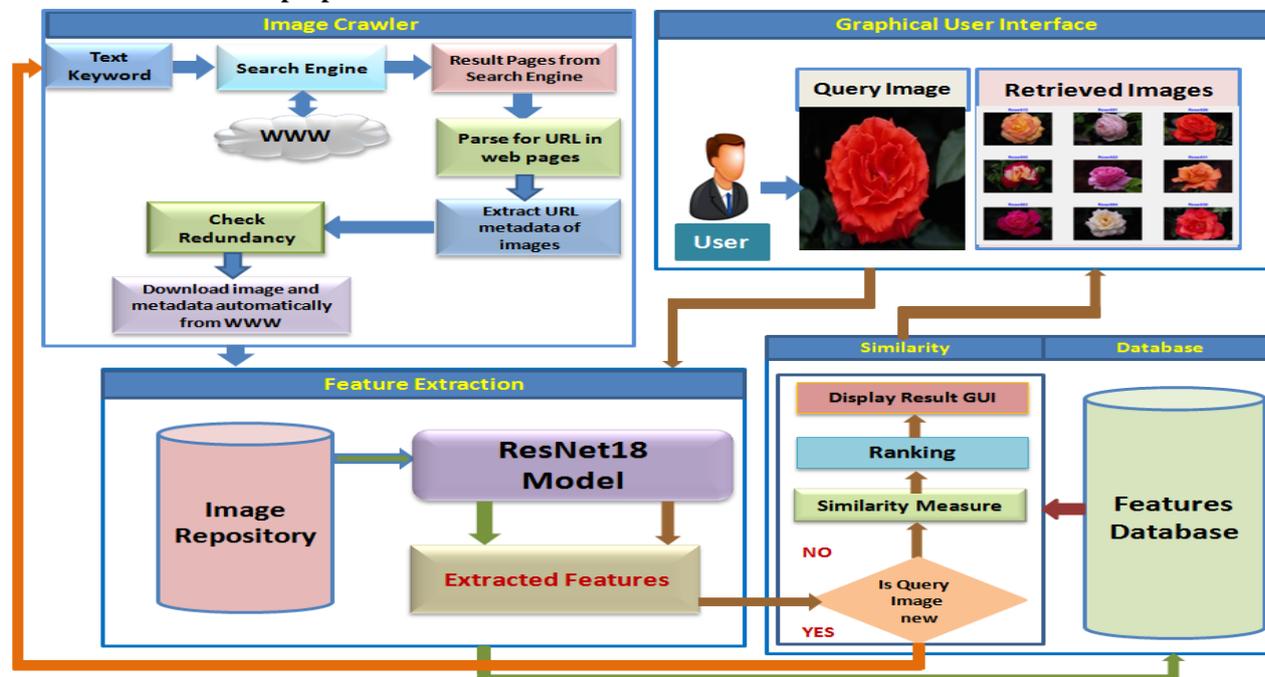


Figure 2. Proposed CBIR Architecture of Distributed Client-Server Model for CBIR System

The architecture of the system shown in figure 2, comprises of the following modules:

- **GUI** – It is client-side module, used for user interaction with the CBIR system. The interface provides the user with input query image by selecting image from the

repository or any image of their choice and submit it. The GUI displays output of the resultant images retrieved by the system.

- **Training and feature extraction** – In this, the ResNet18 model is used to train the images and to extract features. These extracted features are stored in features database for future usage.
- **Database** – Here, all the features extracted by feature extraction module are stored for image retrieval.
- **Similarity** – This module compares features of a query and the features that are already stored in the features database to find the similarity. Later, based on similarity, images are ranked and displayed on the GUI on client side.
- **Web Image Crawler** – This module is invoked whenever a new class of a query image is submitted by the user and an update of the repository is needed.
- **Image repository** – This module is used to store images downloaded from WWW using crawler.

B. ResNet18 model

The Residual Neural Network (ResNet) model was first proposed by Microsoft Research Team [5] and is the winner of ImageNet challenge during 2015. The ResNet is built to overcome the degradation of performance of the network when layers are deeply stacked in the neural network. The single residual block or skip connection used in ResNet is as shown in the figure 3.

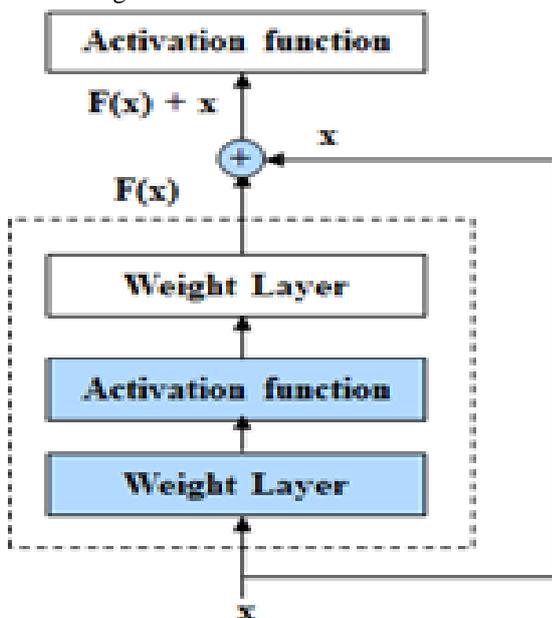


Figure 3 Single residual block or Skipped connections

The skip connections or residual blocks are used to overcome the vanishing gradient problems in deep networks. In the above residual block, the network instead of learning from $x \rightarrow F(x)$, it learns from $x \rightarrow F(x) + x$ mapping, where x is the identity function $G(x)$ and the connection is referred to as shortcut or skip connection. Thus, we have the mapping as $x \rightarrow F(x) + G(x)$, if both input x and output $F(x)$ have same dimensions.

In our proposed work, there are 18 layers used for constructing the ResNet model and hence referred to as ResNet18 as shown in figure 4.

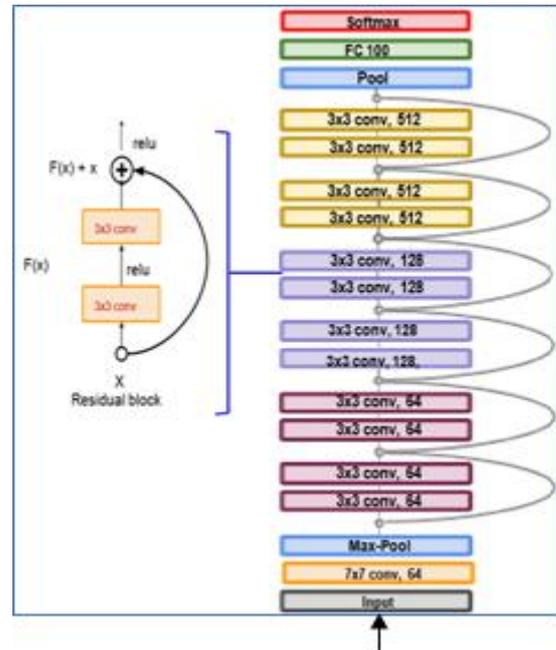


Figure 4 Basic ResNet18 Architecture

The ResNet18 model has the first two layers as 7×7 convolution layer of 64 output channels, with stride as 2 and a max pool layer with 3×3 is used. The next layer is max pool with 3×3 size and with a stride of 2. The batch normalization layer is added after each convolutional layer in the residual block.

The ReLU activation function $y = \max(0, x)$ is used to determine to reduce the unwanted neurons and makes the activations sparse and efficient. This ReLU function has a slope for $x > 0$ else zero for $x \leq 0$.

C. Techniques

The features of a query image are similarly extracted and compared with the features in database. The score obtained for all retrieved images are ranked for similarity and sorted in descending order. These images are displayed on GUI in the client side interface. This model is tested and evaluated for 10,000 images of 50 categories, 20,000 images of 100 categories, 30,000 images of 150 categories, 40,000 images of 200 categories and finally for 50,000 images belonging to 250 categories with sample query images. The algorithm I shows the implementation of the CBIR system using ResNet18 model.

Algorithm I: CBIR using ResNet18 Model

- Step 1: Train the model with 100 categories for 80% images in them and 20% for testing to fix up the weights.
- Step 2: Load the previously trained ResNet18 as feature extractor
- Step 3: Input a query image q
- Step 4: Extract features of q using feature extractor
- Step 5: For every feature r of images in repository calculate Euclidean distance d between q and r
if $d \leq 0.15$ then //85% similarity

Add (imageIndex, d) to resultset

Step 6: Rank results based on d in ascending order

Step 7: Return result to display on GUI

D. Performance Metric

The Euclidean distance is used as similarity metric as given in the equation (6.7).

$$d(q, r) = \sqrt{\sum_{i=1}^n (q_i - r_i)^2}$$

where q is query and r is repository image.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

The following table shows the results obtained in terms of average precision value for number of images ranging from 10,000 to 50,000.

Table 1 Average precision value obtained for varying numbers of images in dataset

Image dataset (size)	Average Precision Value (%)
10,000	98.51
20,000	95.86
30,000	95.10
40,000	96.62
50,000	94.27

Few sample images retrieved using the proposed model is shown in the following figures 5 to 10.

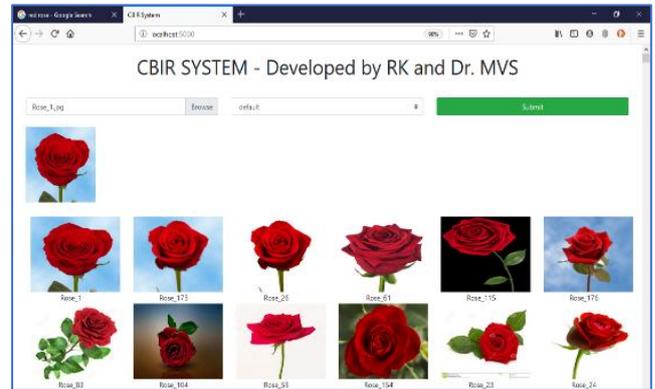


Figure 7 Rose - page 1 results

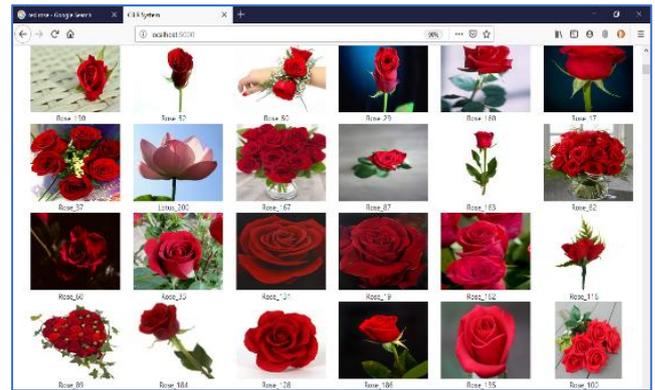


Figure 8 Rose - page 2 results

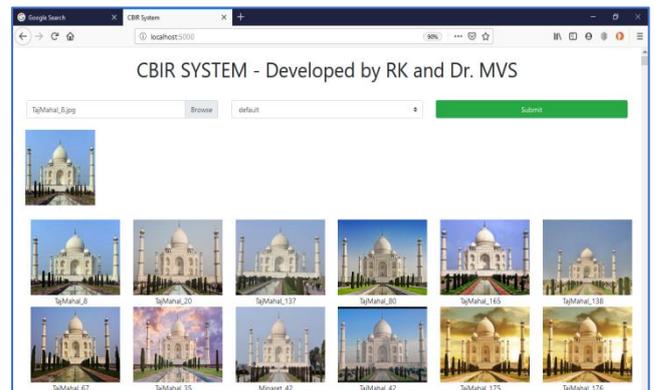


Figure 9 Taj Mahal - page 1 results

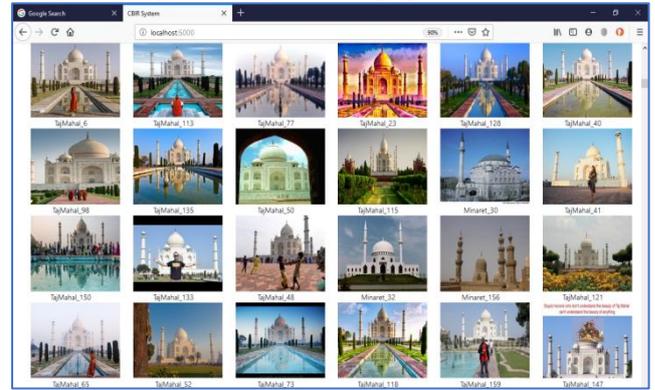


Figure 10 Taj Mahal - page 3 results

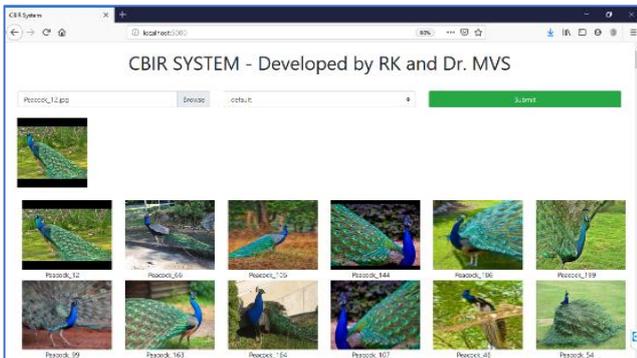


Figure 5 Peacock - page 1 results

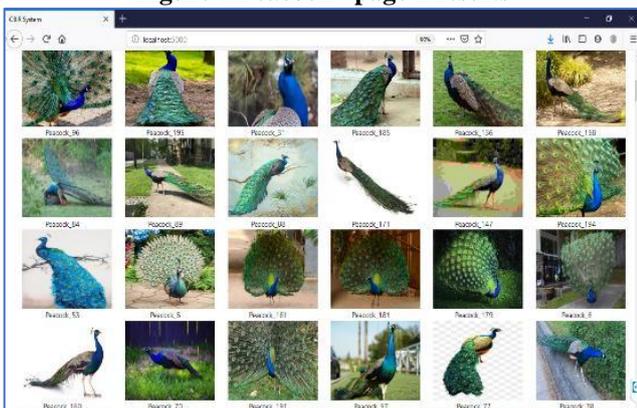


Figure 6 Peacock - page 3 results

The working of crawler module in the CBIR system is invoked automatically whenever a new category of query image is searched. Then, the inbuilt crawler downloads the required images from the web and the time taken for downloading depends upon the Internet bandwidth, size of image, etc. An example of the same is shown in the figures 11 to 14.

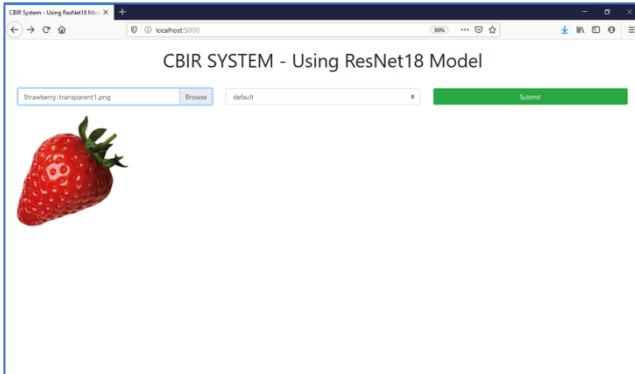


Figure 11 New image category – Strawberry

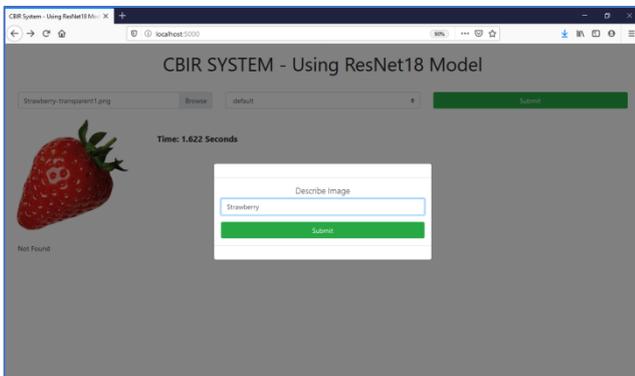


Figure 12 Text query input to download new image

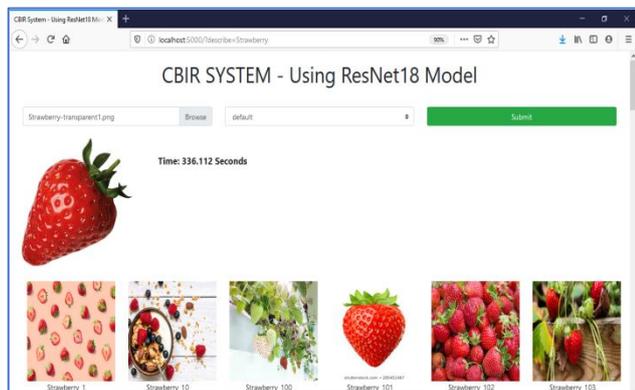


Figure 13 Display of downloaded images – page 1



Figure 14 Display of downloaded images – page 5

After the crawler downloads, the ResNet18 model, extracts the features and stores them for future retrieval process using the CBIR system. The above query image is provided to the CBIR system, then it retrieves similar images and is as shown in the figures 15 to 18.

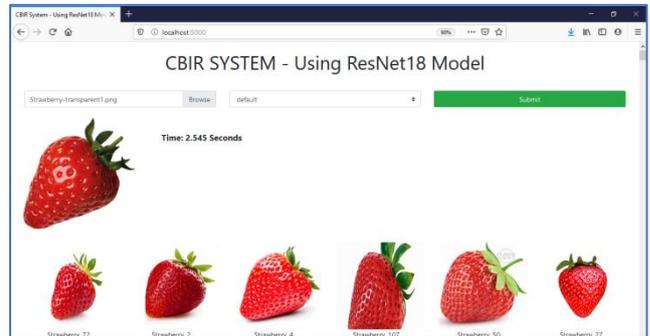


Figure 15 Strawberry result using CBIR system - page 1

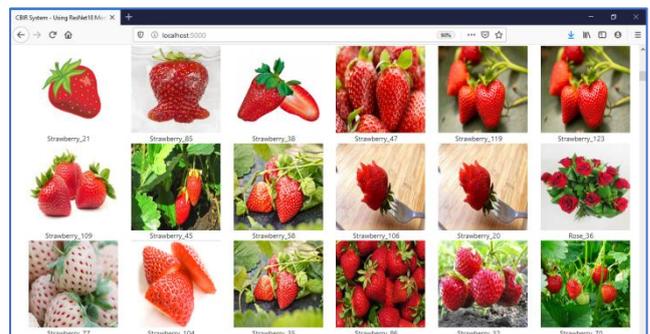


Figure 16 Strawberry result using CBIR system - page 3

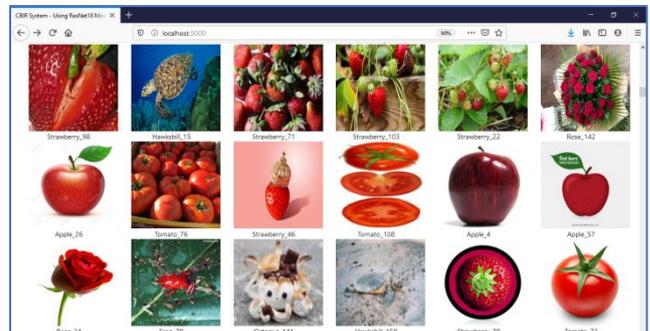


Figure 17 Strawberry result using CBIR system - page 5

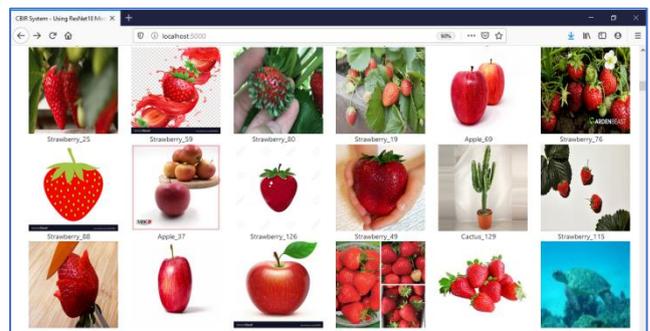


Figure 18 Strawberry result using CBIR system - page 6

The proposed system is evaluated with sample images as query and output values of average precision obtained are tabulated. The Google image search engine uses the CBIR technique and text-based search as well.

A comparison of precision values obtained for sample images that are provided as query images to both the proposed CBIR system and Google's image search engine is shown in figures (19 to 31) below.

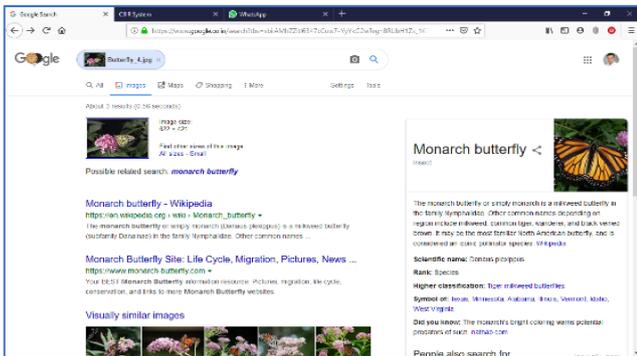


Figure 19 Google search page for Sunflower

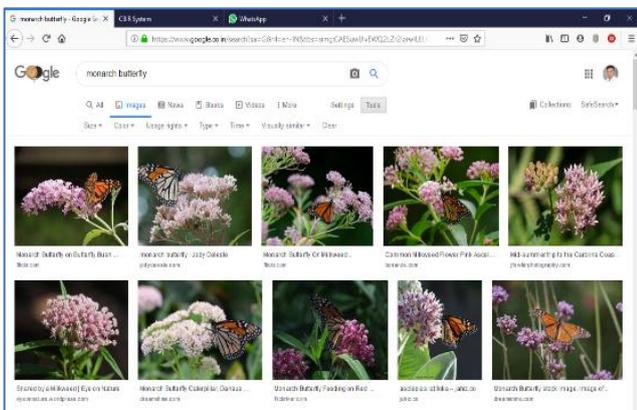


Figure 20 Sunflower result - Google page 1

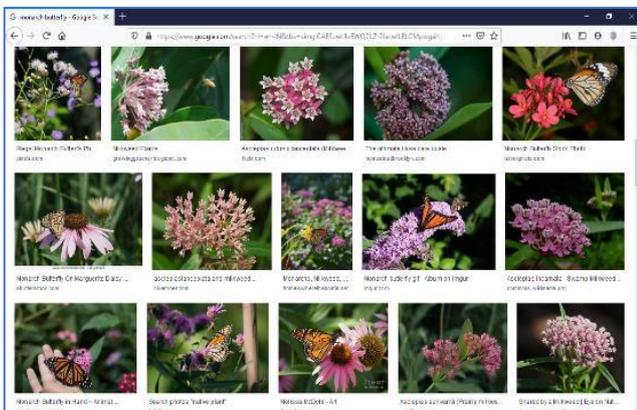


Figure 21 Sunflower result - Google page 3

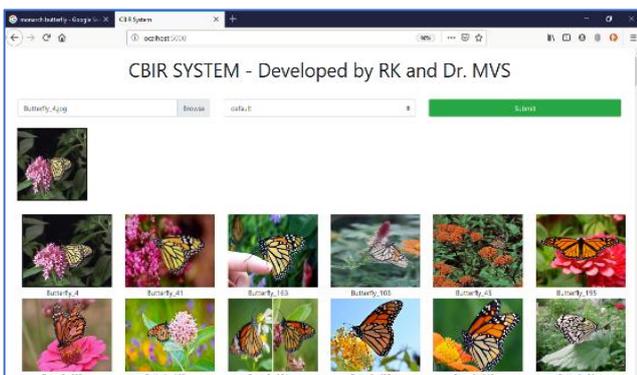


Figure 22 Sunflower result from proposed system - page 1



Figure 23 Sunflower result from proposed system - page 4

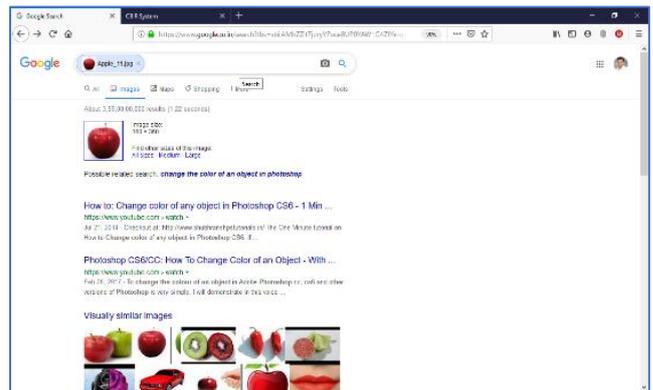


Figure 24 Google search page for Apple

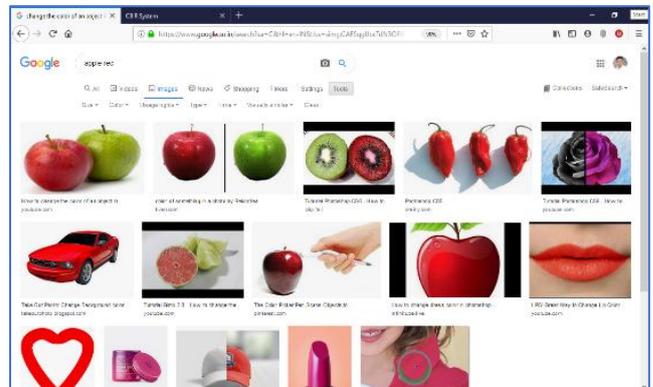


Figure 25 Apple - Google page 1

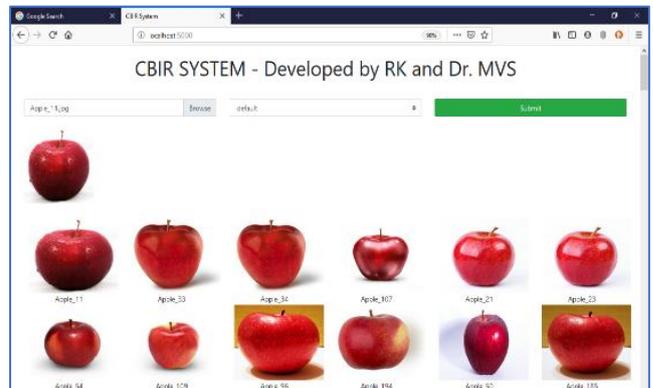


Figure 26 Apple result from proposed system - page 1



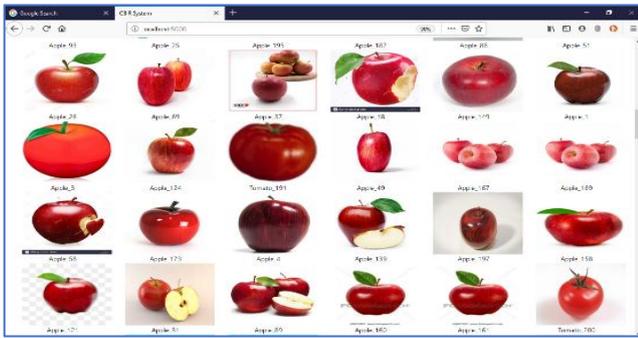


Figure 27 Apple result from proposed system - page 4

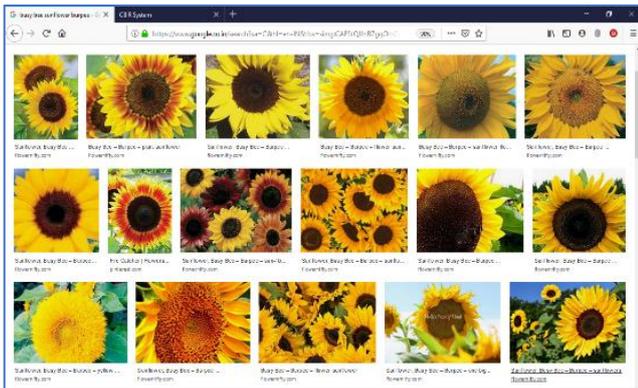


Figure 28 Sunflower - Google page 1

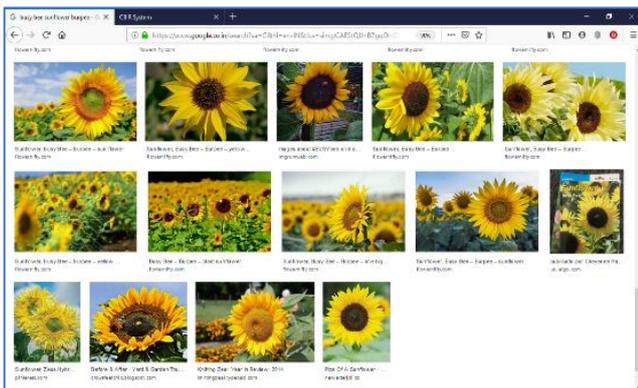


Figure 29 Sunflower - Google page 3

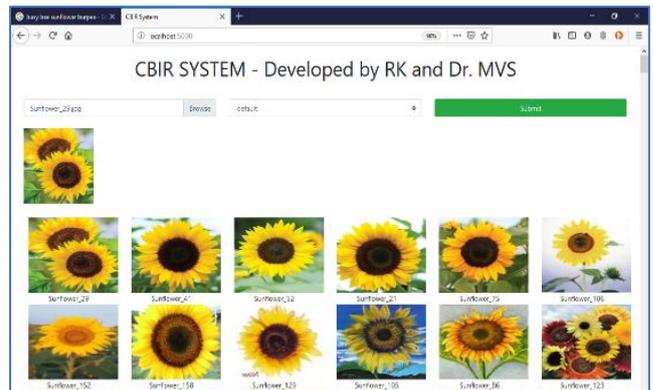


Figure 30 Sunflower result from proposed system - page 1



Figure 31 Sunflower result from proposed system - page 3

Table 2 Average Precision values obtained for Google image search against proposed work

Sl. No.	Category Query	Google Image Search (unknown data size)			Proposed Work (50,000 images)		
		No. of images relevant to query image	No. of images retrieved	Precision %	No. of images relevant to query image	No. of images retrieved	Precision %
1	Red Rose	93	93	100.00	81	81	100.00
2	Horse	78	81	96.29	80	80	100.00
3	Tripod	35	35	100.00	114	114	100.00
4	Swan	71	79	89.87	89	96	92.71
5	Elephants	73	80	91.25	140	144	97.22
6	Sunflower	57	58	98.28	120	120	100.00

7	Taj Mahal	183	183	100.00	150	150	100.00
8	Soccer Ball	160	208	76.92	145	148	97.97
9	Umbrella	255	255	100.00	138	138	100.00
10	Apple	4	15	26.67	71	72	98.61
11	Butterfly	49	102	48.04	88	90	97.78
Average				84.30			98.57

The performance of the proposed system is also compared with the other existing works as in table 3 and graphical representation of the comparison is shown in figure 32. The data size used for the study are also listed.

Table 3 Comparison of existing works and proposed work in terms of average precision value

Works	Data set Size	Average Precision (%)
L.K. Pavithra et. al. [6], 2018	10,000	83.55
Rajkumar et. al. [7], 2019	1,000	84.20
Google Image Search	Unknown	84.30
Guang-Hai et. al. [8], 2019	10,000	86.80
Purohit et. al. [9], 2015	10,000	98.41
Proposed work	50,000	98.57

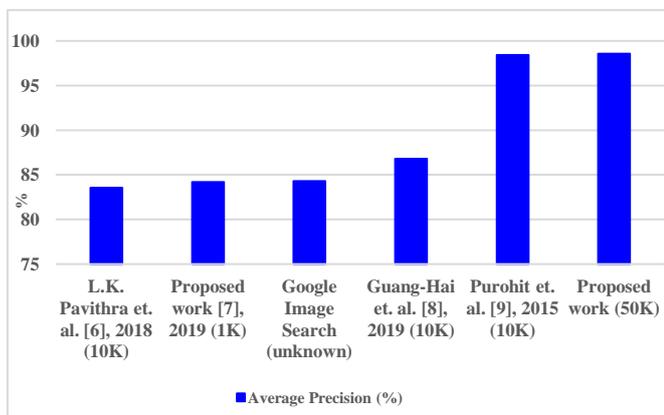


Figure 32 Graphical representation of average precision values of existing works and proposed work

V. CONCLUSIONS AND SCOPE FOR FUTURE WORK

In this paper, the ResNet with the skip connections was used. Here, the proposed system showed a consistent performance for data set ranging from 10,000 to 50,000 images. This proposed system was compared with Google Image Search results for sample query images and had shown performance higher by 14.27% with respect to average precision value. A comparison with existing systems also revealed that our proposed work in terms of average precision value has good performance.

As future work, the experiment can be conducted on different datasets, increase the number of layers and evaluate the performances.

REFERENCES

1. Augusto C. Valente et. al., Comparison of texture retrieval techniques using deep convolutional features, Society for Imaging Science and Technology, <https://doi.org/10.2352/ISSN.2470-1173.2019.8.IMAWM-406> © 2019.
2. Ramesh Babu P, E Sreenivasa Reddy, A Novel Framework design for Semantic based Image retrieval as a Cyber Forensic Tool, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-8 Issue-10, DOI: 10.35940/ijitee.I9593.0881019, August 2019.
3. D. G. Lowe, "Object recognition from local scale-invariant features," Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 1999, pp. 1150-1157 vol.2, 1999.
4. Sivic and Zisserman, "Video Google: a text retrieval approach to object matching in videos," Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 2003, pp. 1470-1477 vol.2.
5. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2016.90, 2015.
6. L.K. Pavithra, T. Sree Sharmila, An efficient framework for image retrieval using color, texture and edge features. Elsevier, Computers and Electrical Engineering 70, 2018, 580–593, 2018.
7. R Rajkumar, Dr. M V Sudhamani, Retrieval of Images using Combination of Features as Color, Color Moments and Hu Moments, Advances in Image and Video Processing, Open Access – UK Journal, DOI: 10.14738/aivp.75.7208, Volume 7 No 5, October (2019); pp: 9-21, 2019.
8. Guang-Hai Liu, Jing-Yu Yang, Zuo Yong Li, Content-based image retrieval using computational visual attention model, <http://dx.doi.org/10.1016/j.patcog.2015.02.005>, Pattern Recognition, Vol 48(2015) 2554–2566, 2015.
9. Purohit Shrinivasacharya, Dr. M V Sudhamani, Extraction of Texture with Wavelet Transforms and HSI Color Features Technique for Content Based Image Retrieval, International Journal of Computer Information Systems, Vol. 12, No.3, 2016.
10. J. Wang and X.-S. Hua. Interactive image search by color map. ACM Transactions on Intelligent Systems and Technology (TIST), Vol. 3, No. 1, Article 12, 2011.
11. Wang, X.-Y., Zhang, B.-B., & Yang, H.-Y, Content-based image retrieval by integrating color and texture features. Multimedia tools and applications, 68(3):545–569, 2012.
12. C. Wengert, M. Douze, and H. Jégou. Bag-of-colors for improved image search. In Proceedings of the 19th ACM international conference on Multimedia, pages 1437–1440. ACM, 2011.
13. Hang Zhang Jia Xue Kristin Dana. Deep TEN: Texture encoding network. In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR), pages 2896–2905. IEEE, 2017.
14. Jinyun Lu, Image Retrieval Based on ResNet and KSH, Advances in Intelligent Systems Research, volume 147, International Conference on Network, Communication, Computer Engineering (NCCE 2018), 2018.
15. Radenović F, Tolias G, Chum O, Fine-tuning CNN image retrieval with no human annotation, IEEE transactions on pattern analysis and machine intelligence, 2018, 41(7): 1655-1668, 2018.
16. Sajja Tulasi Krishna, Hemantha Kumar Kalluri, Deep learning and transfer learning approaches for image classification, International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-7, Issue-5S4, February 2019.



17. Saeid Balaneshin-kordan and Alexander Kotov. 2018. Deep Neural Architecture for Multi-Modal Retrieval based on Joint Embedding Space for Text and Images. In Proceedings of WSDM 2018: The Eleventh ACM International Conference on Web Search and Data Mining (WSDM 2018). ACM, New York, NY, USA, 9 pages, <https://doi.org/10.1145/3159652.3159735>, 2018.
18. Hesamian, M.H., Jia, W., He, X. et al. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Digit Imaging* 32, 582–596 (2019). <https://doi.org/10.1007/s10278-019-00227-x>, 2019.

AUTHORS PROFILE



Mr. R Rajkumar is presently working as Assistant Professor, Dept. of ISE, RNSIT, Bengaluru. has completed B.E and M.Tech degrees in 2001 and 2007 respectively. He has 16 years in teaching, 7 years in research and 2 years industrial experiences. His research area is Content based image retrieval using features and annotations.



Dr. M V Sudhamani, currently working as Dean-R&D, Professor and HoD, Dept. of ISE, RNSIT. She is having Teaching, Research and Industrial experience of 25 years. She has specialization in Image Processing, Content-based Image Retrieval, Advanced Algorithms and Databases. Guided and guiding candidates for Ph. D degree. She has carried out two research projects from VTU and AICTE. She has served as member of Board of Examiners (BOE) and Board of Study (BOS) member in VTU and other autonomous institutions across India. She has organized three international conferences ICDECS 2011, 2015 and 2019.