RS Invariant Image Classification and Retrieval with Pretrained Deep Learning Models

D. N. Hire[1]
E&TC Dept, Research Scholar, DYPCOE, Pune, India

Dr. A. V. Patil[2]
E&TC Dept, Principal, DYPIEMR, Pune, India

***Abstract*—CBIR deals with seeking of related images from large dataset, like Internet is a demanding task. Since last two decades scientists are working in this area in various angles. Deep learning provided state-of-the art result for image categorization and recovery. But pre-trained deep learning models are not strong enough to rotation and scale variations. A technique is proposed in this work to improve the precision and recall of image retrieval. This method concentrates on the extraction of high-level features with rotation and scaling invariant from ResNet18 CNN (Convolutional Neural Network) model. These features used for segregation of images using VGG19 deep learning model. Finally, after classification if the class of given query image is correct, we will get the 100% results for both precision and recall as the ideal requirement of image retrieval technique. Our experimental results shows that not only our proposed technique outstrip current techniques for rotated and scaled query images but also it has preferable results for retrieval time requirements. The performance investigation exhibit that the presented method upgrades the average precision value from 76.50% for combined features DCD (Dominant Color Descriptor), wavelet and curvelet to 99.1% and average recall value from 14.21% to 19.82% for rotated and scaled images utilizing Corel dataset. Also, the average retrieval time required is 1.39 sec, which is lower than existing modern techniques.**

***Keywords*—CBIR; CNN; deep learning; ResNet18; rotation; scale; VGG19**

I. Introduction

Since last two decades due to usage of mobile phones and easily availability of Internet on likely smart devices the data generated per year is increasing day by day. Usually, the data is in the form of images and figures as it is said that "A picture is worth a thousand words". So, the large amount of image data produced and given for various applications like multimedia, e-commerce, medical, agriculture, etc. Managing of that image dataset, for finding, scanning, describing and diving is the tedious task. Managing that large dataset manually may create difficulties like mishandling and wastage of time. In June 2011, Google included the search by image feature as an application to image retrieval [1]. Before that it was search by text method which includes searching of images by its name but as the annotation of every image is difficult task and it varies from ones to others perspective so, the Content Based Image Retrieval (CBIR) invented in 1990s which make use of the low-level description of images like shape, texture, color [2,3] and now a days with deep learning model high level features [4,5] utilized to improve the retrieval result.

From Fig. 1 and 2, it proved that if the image rotated by 90-degree and given as query image, Google can't retrieve similar correct similar images. Google treating image of building as horizontal cylindrical rods and providing images as per this prediction is incorrect. So, there is a need for geometrically transformed efficient image retrieval method. So, for summarizing, features used for image comparison in these current methods are not robust to rotation and scaling factor.

In this paper, we proposed a technique of image retrieval with deep learning which is robust to rotation and scaling by introducing rotated and scaled images in training dataset. And for extracting high-level features from images ResNet18 model utilized whereas for classification VGG19 used.

The remaining part of paper arranged as mentioned: Section II describes the similar work on this research work. Section III provides information related to VGG19 and ResNet18 models used in proposed technique. The proposed RS invariant image retrieval technique explained in Section IV. The proposed method investigation results and discussion mentioned in Section V. The article concluded with conclusion in Section VI.

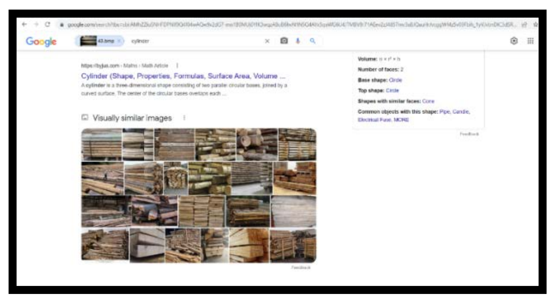

Fig. 1. Result of Google Search for Corel Dataset Image under Category Building with 90-Degree Rotation.

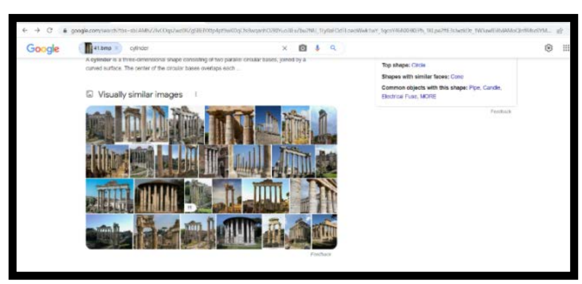

Fig. 2. Result of Google Search for Corel Dataset Original Image under Category Building.

II. Similar Work

The heart of CBIR is representation of image with its features. In that manner the feature extraction process plays vital role in image retrieval. Low level features were considered at the beginning of CBIR system, viz., color, texture, shape, etc. Out of which color is the most prominent feature which will be considered as invariant to scale and rotation. Many researchers already worked on that feature and came up with different techniques as mentioned in [14]. The combination of features like texture and color provided in paper [18] further improved the performance of image retrieval by developing method called RSHD. In [6][7] author collected color and texture features from images with color histogram and Gabor filter algorithm and selected only prominent features for retrieving similar images to reduce the computational complexity. In [8] color and texture feature called CCM and DBPSP respectively and third feature called based on color histogram called CHKM used for image retrieval. [12][13][14] [17] describes the use DCD features in various means to improve accuracy of image retrieval. Use of curvelet features also provided improvement in accuracy of image retrieval as referenced in [15][16][17]. Wavelet features applied in image retrieval as reflected in [9][10][11][17]. Deep Learning technique evolved in 1943 in the form of computer model developed by Walter Pitts and Warren McCulloch using neural networks derived from the human brain. But the deep learning usage started in real manner after the invention of high-speed computer and GPUs to fulfill the memory requirement of deep learning models. In [4][5] deep learning model provided state-of-the-art result in case of image recognition for large image dataset containing approximately 2.5 million images. So, there is a scope of research in this field of deep learning to further improve image retrieval performance by considering rotation and scale invariant features.

III. Deep Learning Models

There are various deep learning models invented out of which CNN prominently used for image related applications. Using CNN models various algorithms developed out of that we choose ResNet18 and VGG19 by comparing performance related to computational complexity and accuracy.

A. *ResNet18*

ResNet-18 is an 18-layer deep convolutional neural community. The community can classify snap shots into 1001 kind of item categories, consisting of keyboards, mice, pencils, and a number of animals. As a result, the community has found out a number of wealthy characteristic representations for a number of images. The community's image enter length is 224 × 224 pixels. We can do the modifications in network by transfer learning as per our requirement.

ResNet18 model is used to create bag of features in our proposed technique due to its advantageous properties.

- Networks with a large number of layers (even thousands) may be easily taught without raising the training error rate.
- ResNets can aid with identity mapping to solve the vanishing gradient problem.

Fig. 3 describes the working of ResNet18 model. The output of a specific layer is linearized and used as a feature vector given the network. We experiment with two alternative layers of the network: the output of the average pooling layer and the linearized output of the fifth convolutional stage (that is conv5x) (that is average pool). The size of the feature vector for the conv5x layer is 25,088 (77512), and for the average pool layer, it is 512. Since the feature vector's size influences computation costs, dimensionality reduction techniques like Principal Component Analysis are used to lower the feature vector's size.

Layer Name	Output Size	ResNet-18
conv1	$112 \times 112 \times 64$	7×7, 64, stride 2
conv2_x	$56 \times 56 \times 64$	3×3 max pool, stride 2
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
conv3_x	$28 \times 28 \times 128$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
conv4_x	$14 \times 14 \times 256$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
conv5_x	$7 \times 7 \times 512$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
average pool	$1 \times 1 \times 512$	7×7 average pool
fully connected	1000	512×1000 fully connections
softmax	1000	

Fig. 3. ResNet18 Architecture [19].

B. *VGG19*

This network shown in Fig. 4, received a fixed size RGB image (224 * 224) as input, indicating that the matrix has shape (224,224.3). The only pre-processing is to subtract the average RGB value of each pixel, which was calculated throughout the training set. They used kernel with size (3*3) and step size of 1 pixel to cover the whole concept of image. To maintain the spatial resolution of the image, spatial padding was applied on 2*2 pixel window. Then, the adjusted linear unit (ReLu) is used to introduce nonlinearity into the model to improve classification and save processing time (previous models used tanh or sigmoid). Three fully connected layers have been implemented, the first two having size 4096, followed by a 1000 channel layer for 1000 line ILSVRC classifier and finally a softmax function.

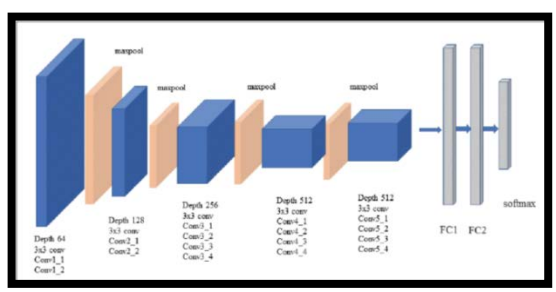

Fig. 4. VGG19 Architecture [20].

IV. RS INVARIANT IMAGE RETRIEVAL SYSTEM

To make the system Rotation and scale (RS) invariant we trained the VGG19 model of classification with images containing rotated images with 45-degree and 90-degree and scaled images with scaling factor 0.4 and 0.6. So, the Corel-1k original dataset which consist of 10 classes and in each class 100 images now transformed to 5k dataset as each image in 1k dataset is rotated with 45-degree, 90-degree and scaled with scaling factor 0.4 and 0.6. In that way 1 image in 1k dataset will have another 4 images. Fig. 5 illustrates the detailed working model of proposed system. Firstly, the dataset divided into training and testing parts. Then training dataset trained with ResNet18 model and features of that dataset collected from pooling layer 5 of this model. Then these collected features acts as the input to the VGG19 model and used for classification of query images to find the class of given query image.

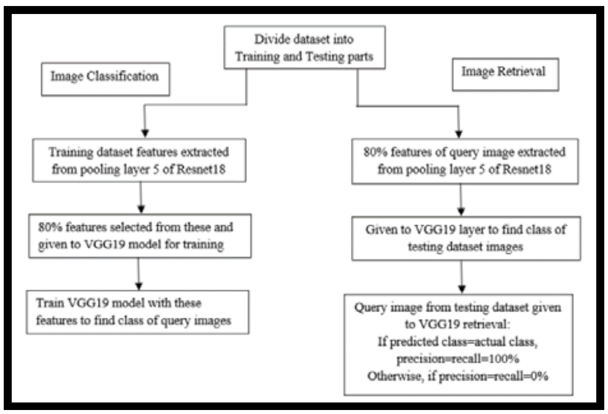

Fig. 5. The Proposed RS Invariant Image Retrieval System.

In the image retrieval part, 80% features of query image extracted from pooling layer 5 of ResNet18 model and given to VGG19 model to find class of query image. Then after classification for output class top 12 images will be retrieved at the output. If predicted class is same as actual class of query image then precision and recall parameter values will be 100% but if predicted class is wrong then precision and recall values goes down to 0%.

V. EXPERIMENTAL STRUCTURE AND RESULTS

To visualize the practicality of our proposed system analyzed using datasets corel-1k, corel-1k scale and rotate dataset with 45-degree and 90-degree rotated and scaled images including images with scaling factor 0.4 and 0.6, Corel-1k rotated dataset which includes images rotated with angle 0-degree, 90-degree, 180-degree and 270-degree as well as Corel-1k scaled dataset which consists of images scaled with factor 0.5, 0.75, 1, 1.25, and 1.5.

Table I describes the dataset with their categories, classes, total images and total images per class in detail. VGG19 model trained for classification of images in dataset. The dataset divided into training and testing dataset as 70% and 30% respectively. The classification accuracy for every dataset is given in Table II. The classification accuracy is for all dataset is more than 96% which is providing better results for image retrieval.

TABLE I. DETAILS OF DATASET USED

Name of dataset	Description	Classes	Total images	Images per class
Corel-1k	Original dataset	10	1000	100
Corel-1k Scale	Scaled by factor 0.5, 0.75, 1, 1.25, 1.5	10	5000	500
Corel-1k rotate	Rotated by degree-0, 90, 180, 270	10	4000	400
Corel -1k scale & rotate	Scaled by factor 0.4, 0.6 and rotated by angle 45 & 90	10	5000	500

TABLE II. CLASSIFICATION ACCURACY OF SELECTED DATASET

Dataset	No. of Training Images	No. of Testing Images	Classification accuracy (%)
Corel-1k	700	300	97
Corel-1k Scale	3500	1500	98.3
Corel-1k rotate	2800	1200	96.8
Corel -1k scale & rotate	3500	1500	96.5

Fig. 6 implies the confusion matrix for Corel-1k scale dataset. After analysing this, our model gives 100% classification accuracy for 5 classes bus, dinosaurs, elephant, food and horse whereas minimum accuracy for class building in which out of 150 testing images, 140 classified correctly but 5 classified as African people and 5 as beach due to the features extracted from images.

Correctly classified and retrieved images for query image from Corel-1k rotate dataset and classAfrican People illustrated in Fig. 7. Fig. 8 shows the evidence of wrong classification of query image which actually belongs to African people category classified under elephant category and resulting in 0% precision and recall values.

Fig. 9 and 10 demonstrate the analysis of proposed technique in the form of precision and recall for all query images from dataset Corel-1k, which is highest in all the previous CBIR methods. For the proposed technique the precision values lies between 99% and 100%, whereas recall values lies between 19% and 20% (the maximum value of recall is 20% as retrieving top 20 images).

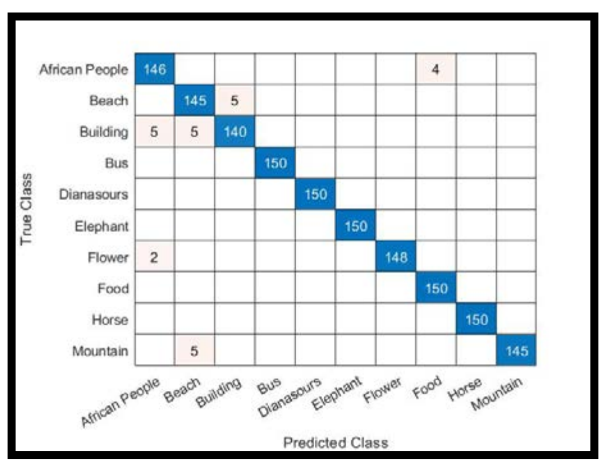

Fig. 6. Confusion Matrix for Corel-1k Scale Dataset.

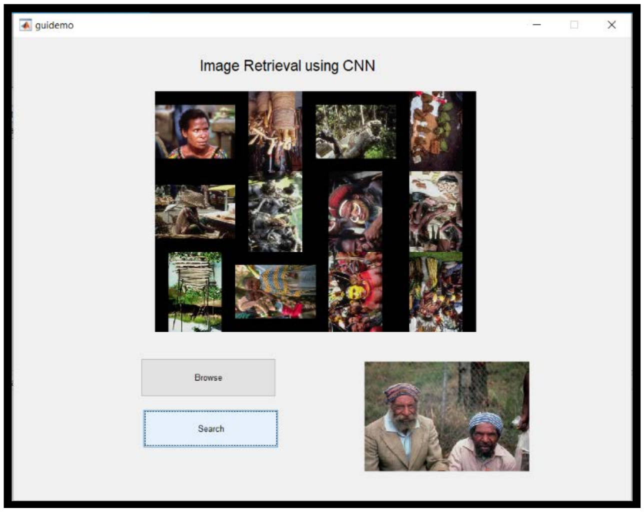

Fig. 7. Correctly Classified & Retrieved Image of Category ‘African People’ from Corel-1k Rotate Dataset.

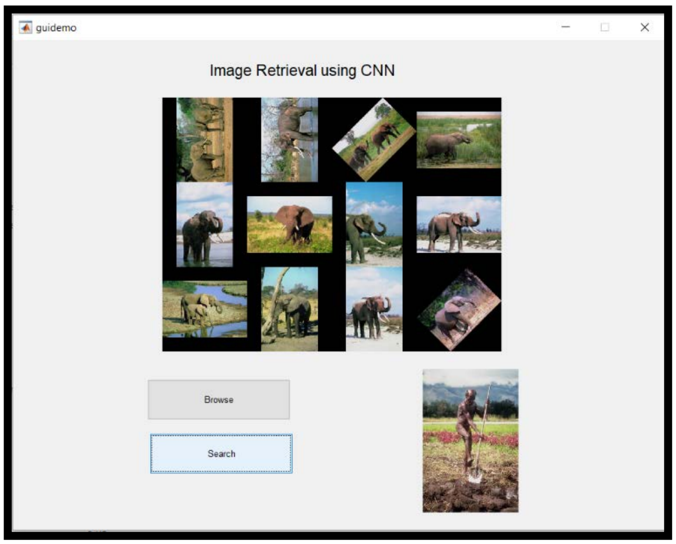

Fig. 8. Wrongly Classified & Retrieved Image of Category ‘African People’ from Corel-1k Scale & Rotate Dataset.

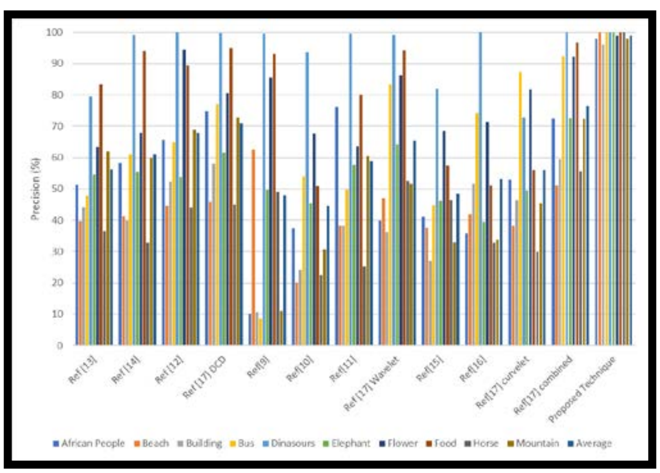

Fig. 9. Comparison of Proposed Technique % Precision with Previous Techniques using Corel-1k Dataset.

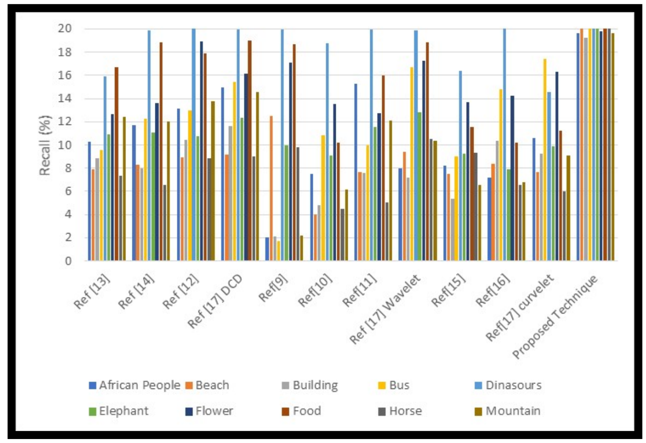

Fig. 10. Comparison of Proposed Technique % Recall with Previous Techniques using Corel-1k Dataset.

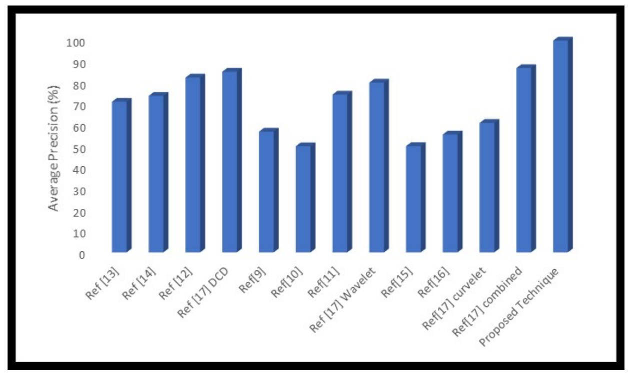

Fig. 11. Comparison of Proposed Technique % Precision with Previous Techniques using Corel-1k scale Dataset.

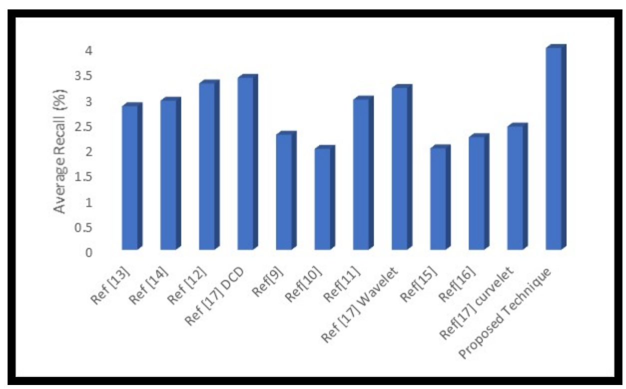

Fig. 12. Comparison of Proposed Technique % Recall with Previous Techniques using Corel-1k scale Dataset.

In case of Corel-1k scale dataset comparison for parameters precision and recall with previous techniques proves our system as scale invariant image retrieval as illustrated in Fig. 11 and 12. Fig. 13 shows the performance of proposed system for various datasets. The system gives best result for Corel-1k scale dataset which shows robustness of system for scale invariance. Average retrieval time for proposed system calculated by randomly considering 100 query images and averaging their retrieval time and plotted in Fig. 14 which shows that the retrieval time is less for proposed system except retrieval time of system mentioned in [18].

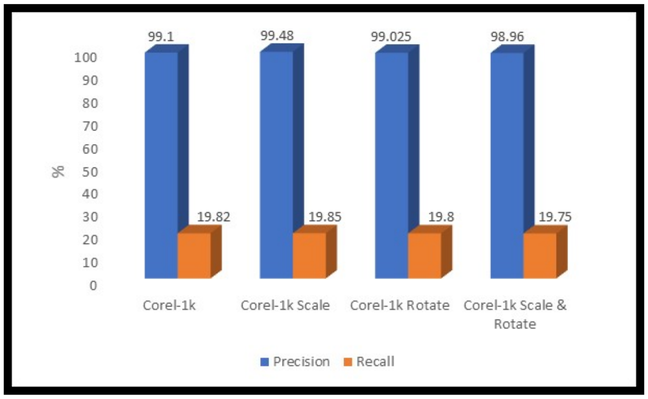

Fig. 13. Performance of Proposed Technique for Selected Dataset.

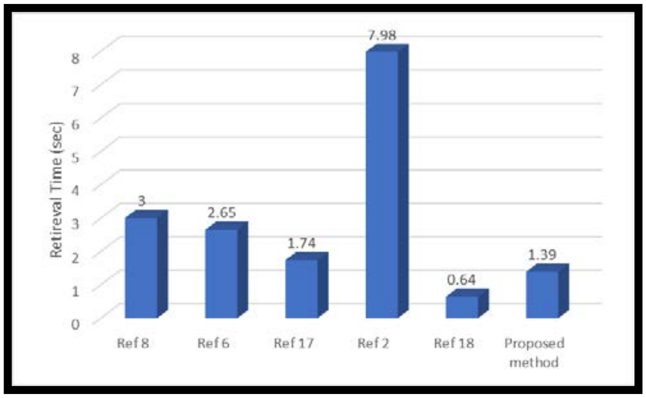

Fig. 14. Average Retrieval Time of Image Retrieval Systems.

VI. Conclusion

We present a deep learning-based feature extraction, classification, and retrieval approach that is resistant to scale and rotation fluctuations in this paper. The most important 80% of collected features are chosen, reducing the computational complexity of the classification and retrieval procedure. In comparison to existing state-of-the-art systems, the experimental findings demonstrated that the suggested CBIR system had the highest precision and recall rate for all datasets, including rotated and scaled images. In addition, compared to previous approaches, there is a lower demand for average retrieval time. Because it has to display the images at output from a selected class rather than the entire database, the classification procedure before retrieval reduces image retrieval time. When the query image is correctly classified, retrieval accuracy is 100 percent.

Acknowledgment

This work supported by Dr. D. Y. Patil Institute of Engineering, Management & Research, Akurdi, Pune by providing access to latest research articles.

References

[1] Accessd on 19th February, 2022 https://en.wikipedia.org/wiki_Google_Images.

[2] Kundu, Malay Kumar, Manish Chowdhury, and Samuel Rota Bulo, "A graph-based relevance feedback mechanism in content-based image retrieval", Knowledge-Based Systems, 2015, 73, pp. 254–264.

[3] Yue, Jun, Zhenbo Li, Lu Liu, and Zetian Fu.: '"ontent-based image retrieval using color and texture fused features", Mathematical and Computer Modelling, 2011, 54, (3), pp. 1121–1127.

[4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[6] ElAlami, M. Esmel., "A novel image retrieval model based on the most relevant features", Knowledge-Based Systems, 2011, 24, (1), pp. 23–32.

[7] Wang, Xiang-Yang, Hong-Ying Yang, and Dong-Ming Li., "A new content-based image retrieval technique using color and texture information", Computers & Electrical Engineering, 2013, 39, (3), pp. 746–761.

[8] Chuen-Horng Lin, Rong-Tai Chen and Yung-Kuan Chan, "A smart content-based image re- trieval system based on color and texture feature", Image and Vision Computing, 2009, **27**, (6),pp. 658–665.

[9] Farsi, Hassan, and Sajad Mohamadzadeh, "Colour and texture feature-based image retrieval by using Hadamard matrix in discrete wavelet transform", IET Image Processing, 2013, 7, (3), pp. 212–218.

[10] Shiv Ram Dubey, Satish Kumar Singh and Rajat Kumar Singh, "Local Wavelet Pattern: A New Feature Descriptor for Image Retrieval in Medical CT Databases", IEEE Transaction on Image Processing, 2015, 24, (12), pp. 5892–5903.

[11] Murala, Subrahmanyam, Anil Balaji Gonde, and Rudra Prakash Maheshwari, "Color and texture features for image indexing and retrieval", IEEE International Advance Computing Conference (IACC), 2009, pp. 1411–1416.

[12] Dubey, Shiv Ram, Satish Kumar Singh, and Rajat Kumar Singh, "Local neighbourhood-based robust colour occurrence descriptor for colour image retrieval", IET Image Processing, 2015, 9, (7), pp. 578–586.

[13] Talib, Ahmed, Massudi Mahmuddin, Husniza Husni, and Loay E. George, "A weighted dominant color descriptor for content-based image retrieval", Journal of Visual Communication and Image Representation, 2013, 24, (3), pp. 345–360.

[14] Deng, Y., Manjunath, B. S., Kenney, C., Moore, M. S., & Shin, H., "An efficient color representation for image retrieval", IEEE Transactions on Image Processing, 2001, 10, (1), pp. 140–147.

[15] Sumana, I. J., Islam, M. M., Zhang, D., and Lu, G., "Content based image retrieval using curvelet transform", 10th Workshop on IEEE Multimedia Signal Processing, Cairns, Qld, 2008, pp. 11–16.

[16] Anil Balaji Gonde, R.P. Maheshwari and R. Balasubramanian, "Modified curvelet transform with vocabulary tree for content based image retrieval", Digital Signal Processing, 2013, 23, (1), pp. 142–150.

[17] Sadegh Fadaei, Rassoul Amirfattahi, Mohammad Reza Ahmadzadeh, "A New Content-Based Image Retrieval System Based on Optimized Integration of DCD, Wavelet and Curvelet Features" Feb-2017, IET Image Processing, Vol 11, issue 2, 89-98.

[18] Shiv Ram Dubey, Satish Kumar Singh, and Rajat Kumar Singh, "Rotation and scale invariant hybrid image descriptor and retrieval", Computers & Electrical Engineering, 2015, 46, pp. 288–302.

[19] Paolo Napoletano, Flavio Piccoli, Raimondo Schettini, "Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity", accessd on 19th February, 2022 https://www.researchgate.net/figure/ResNet-18 Architecture_tbl1_322476121.

[20] Roland Hewage, "Extract Features, Visualize Filters and Feature Maps in VGG16 and VGG19 CNN Models", accessd on 19th February, 2022 https://towardsdatascience.com/extract-features-visualize-filters-and-feature-maps-in-vgg16-and-vgg19-cnn-models-d2da6333edd0.