International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
***Article Received:** 30 July 2023 **Revised:** 18 September 2023 **Accepted:** 02 October 2023*

Image Retrieval Using Auto Encoding Features In Deep Learning

Syed Qamrul Kazmi[1], Munindra Kumar Singh[2], Saurabh Pal[3]
[1]Research Scholar, Department Of Computer Applications
V.B.S. Purvanchal University
Jaunpur, U.P., INDIA
syed.qamrulkazmi@gmail.com
[2]Assistant Professor, Department Of Computer Applications
V.B.S. Purvanchal University
Jaunpur, U.P., INDIA
munindra09_vbspu@yahoo.in
[3]Professor, Department Of Computer Applications
V.B.S. Purvanchal University
Jaunpur, U.P., INDIA
drsaurabhpal@yahoo.co.in

Abstract— The latest technologies and growth in availability of image storage in day to day life has made a vast storage place for the images in the database. Several devices which help in capturing the image contribute to a huge repository of images. Keeping in mind the daily input in the database, one must think of retrieving those images according to certain criteria mentioned. Several techniques such as shape of the object, Discrete Wavelet transform (DWT), texture features etc. were used in determining the type of image and classifying them. Segmentation also plays a vital role in image retrieval but the robustness is lacking in most of the cases. The process of retrieval mainly depends on the special characteristics possessed by an image rather than the whole image. Two types of image retrieval can be seen. One with a general object and the other which may be specific to some type of application. Modern deep neural networks for unsupervised feature learning like Deep Autoencoder (AE) learn embedded representations by stacking layers on top of each other. These learnt embedded-representations, however, may degrade as the AE network deepens due to vanishing gradient, resulting in decreased performance. We have introduced here the ResNet Autoencoder (RAE) and its convolutional version (C-RAE) for unsupervised feature based learning. The proposed model is tested on three distinct databases Corel1K, Cifar-10, Cifar-100 which differ in size. The presented algorithm have significantly reduced computation time and provided very high image retrieval levels of accuracy.

Keywords- Image Retrieval, Content based image retrieval, Auto encoding, Feature extraction

I. INTRODUCTION

Content Based Image Retrieval is a technique which is used to search similar type of images in a large database. The standard procedure applied to this technique is to find images by extracting certain features from images. The results we get are very helpful when we are primarily interested in finding similarity in a query image and similar matching results across the database. The information content must describe the features that are present in a particular image. For this we have the low-level features and the high-level features categorizing the content which may help in efficient retrieval for a given query image.

The CBIR in its initial stages started in the year 1995 when IBM introduced the first commercially available CBIR system, commonly referred to as QBIC or Query By Image Content [26], which enables users to search for data using user-generated sketches, example photographs, and drawings. The system utilizes textures, shapes, and colors. The Color Co-occurrence Matrix, commonly known as CCM is an information extraction method commonly used in CBIR studies [34] using low level features. In Bose et al [28] they have extracted features from MPEG-7 standard [18, 21] using visual descriptors as well as CCM characteristics, with some success. Lohite et al. [25] used Support Vector Machine (SVM) classifier to optimize the obtained results in terms of Texture, edge and color features. Mehmood et al [27] , in their paper discussed about WATH (weighted average of triangular histograms) of visual words which is another CBIR method that can be used efficiently. A weighted average of triangular histograms is a mathematical operation used in various fields, including image processing and data analysis. To calculate a weighted average of triangular histograms, you typically follow steps which include Triangular histograms, assign weights, normalize histograms and then calculate weighted average.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
***Article Received:** 30 July 2023 **Revised:** 18 September 2023 **Accepted:** 02 October 2023*

The new scheme proposed by Rashno et al [21] proposes to convert the supplied image which is generally in RGB format into three neutrosophic domain sub-segments. A statistic component, a histogram, and color characteristics (including dominant color descriptor (DCD)) are recovered for each sub-segment, which are then utilized to extract similar images.

Real-world circumstances present difficulties in coping with enormous amounts of unlabeled data. The manual labelling procedure is labor-intensive, costly, and necessitates the knowledge of the numbers [12]. Furthermore, supervised feature learning may introduce biases by depending solely on labelled data in addition to being unable to benefit from unlabeled data. Feature learning (or feature extraction) based on unsupervised deep learning has so attracted a lot of attention. Although while deep learning has emerged as the dominant technique with cutting-edge performance in many fields, it suffers from the vanishing gradient problem, which means that as networks go deeper, their performance becomes saturated or even starts to decline quickly. [13]. As a result, shallow counterparts are capable of outperforming deep networks [17]. To solve the performance degradation issue, It was proposed by He et al. that residual blocks should be included between layers [17]. These networks are referred to as ResNets [18]–[22]. Despite the existence of the idea of incorporating residual connections, there has been limited research into the application of this concept to unsupervised segment learning. Furthermore, the current research does not address the issue of performance degradation of deep neural networks when learning unsupervised features. To address this issue, a system has been developed to facilitate the unsupervised learning of features. using residual blocks in AE architectures [22] in this study. While constructing dimensionality reduction experiments, it can be difficult to determine the ideal number of hidden layers, especially for unlabeled data. Even with more layers, the suggested technique will always perform similarly or better. Users benefit from being able to design fewer experiments with large networks since they are confident that the performance of dimension reduction on the network won't be negatively impacted. We must demonstrate that RAEs suffer less from unsupervised feature learning performance deterioration than AEs as the depth of the networks is increased in order to prove our hypothesis. We contrast the proposed method with two important categories of methodologies The literature suggests that autoencoders are the most widely used deep learning architecture for reducing unsupervised dimensions. To evaluate the effectiveness of autoencoders in "feature learning", the same models were evaluated without and with residual connections. This was done to compare the results of seven different methods to determine how autoencoders enhance feature learning compared to other versions of AE in the literature. Standard Autoencoder is the focus of this study, as it is the most widely used. The following is how the paper is structured: The literature study is presented in Section 2, the Resnet Auto Encoder architecture is presented in Section 3, the experiments and results are shown in Section 4, and the article's conclusions are presented in Section 5.

II. RELATED WORK

Afshan et al [15] in their research tried to link the mage feature recognition and visual characteristics attained by humans. The aim was to gain a comprehensive understanding of the most recent developments in Content Based Image Retrieval and image representation. Starting from basic feature extraction to more advanced techniques in semantic deep learning, the research focused on the key features of various image retrieval and image representation models.

Jiaohui Yu [12] widely discussed that the difficulties posed by image retrieval systems are due to the vastness and number of features employed, which can span from hundreds of dimensions to thousands. This phenomenon has been dubbed the dimensionality disaster. Researchers have proposed a range of approaches based on approximation, however, multifactorial image retrieval techniques involving partial differential equations are more commonly employed in daily life.

Himani Chugh and Sheifali Gupta [11] emphasized that it is important to note that the image being queried is located within the dataset, and the Color Difference Histogram (CDH) descriptor is utilized to access images within the database. The CDH function quantifies the difference in color between two labels in the laboratory color space.

Jaya H. Dewan and Sudeep D. Thepade [10] discussed that the purpose of content-based image retrieval is to extract the low-level elements of an image (e.g. colors, textures, shape features, etc.) and compare the query’s image feature vector with the dataset’s images feature vectors to acquire similar images. This approach has been highly sought after by researchers due to its use of visual elements. The purpose of this article is to assess various image retrieval techniques that focus on extraction of features, description of content, and comparison of content.

Shukran et al [7] in their paper examined the various cognitive behavior modification (CBIR) techniques classified as color-based, texture-based, shape-based, and hybrid-based. It also provides a comparative analysis of color-based features, texture-based features, and hybrid techniques, using multiple parameters such as precision, recall, and response time..

Yu Zhao [8] focused on the statistical feature approach of the double-Tree complex wavelet was initially developed to recover edge information in document images by combining the statistical characteristic method with the human eye's visual features. On this basis, meaningful texture features were

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
***Article Received:** 30 July 2023 **Revised:** 18 September 2023 **Accepted:** 02 October 2023*

identified and texture descriptors were used to define document image characteristics. The content elements of the image are integrated into the image automatically, with the description used as a reference, and appropriate comparison assessment criteria are employed to ensure efficient retrieval. This technique has been demonstrated to be highly efficient in terms of retrieval efficiency, while also reducing the complexity of traditional document image retrieval algorithms, as demonstrated by the results.

III. PROPOSED METHODOLOGY

In general CBIR approach, an image is represented by a collection of low-level or high-level elements. Feature encoding, on the other hand, is the practice of transforming an image into a n-dimensional image feature vector, either from RGB or from HSV space. For the purpose of this research, it is suggested to employ a few predefined deep learning models in order to extract image feature vectors. To begin, a brief introduction is provided to the fundamental concepts such as neural networks, predefined models, etc.

3.1 Data Mining and the Traffic Issue

The use of data mining technologies in transportation networks is not very common. In fact, the development of system detector data collecting with the emergence of ITS is very young and has not yet been fully utilized. Although it may be thought of as unpredictable and unmanageable, traffic can actually be shown to be somewhat predictable with the use of archived data, and control can be enhanced with its use. Other DOTs have investigated cutting-edge control methods such traffic responsive and second generation, where the system detector typically composed of three different types of layers, data is required to enable such control approaches, but it has not been discovered that they are used for TOD signal control. Finding patterns in data and creating classifications are made possible by data mining technologies, and the transportation industry can greatly benefit from these ideas. These data mining approaches have been applied in numerous other domains and fields to obtain comparable outcomes from a variety of data sets.

3.2 Neural Network

Neural networks consist of a network composed of non-cyclical arrays of neurons connected to each other. These networks are often depicted as a network composed of layers. The linked nodes known as neurons or artificial neurons are grouped into layers, with each layer in charge of a different sort of processing. A neural network's layers are classified into three types as input layer, hidden layers and output layer. The following examples illustrate a fully connected neural network [11].

3.3 Convolutional Neural Network

A Convolutional Neural Network (CNN) is a type of neural network that accepts images as inputs and manages architecture in a more intelligent manner. Each layer of a CNN consists of neurons arranged in a three-dimensional pattern, with the third dimension being the activation volume. Unlike a Fully-Connected Neural Network, each layer of a CNN is connected to a restricted portion of the previous layer. As discussed previously, a Simple CNN [3] is composed of layers, each layer of which can be transformed into a different level of activation by using differentiable functions. CNN designs are
namely Fully Connected layers, convolutional layers and pooling layers.

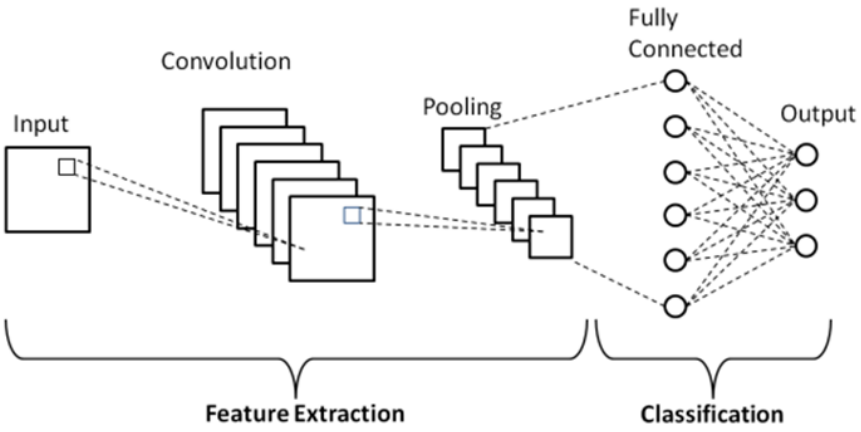

Figure 1: Basic CNN architecture in the image preprocessing stage

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

3.4 Image Acquisition

Firstly in image processing there is generally the capturing of Image by the means of a digital camera in order to transform the real world image to a digital one. It is then analyzed, may be transformed and can be reproduced for different tasks. The image that is acquired is usually raw in nature and it has no significance [20] until it is processed to a meaningful form. An operation involving image processing typically begins with image acquisition. Once the image has been captured, a variety of operations [5] can be performed on it and the operations related to a specific image can be completed. However, further tasks that will be processed successfully can only be accomplished if the image has been accurately captured and acquired. Furthermore, the image enhancement methods [6] will not be successful if the image has not been captured accurately. Consequently, many successful results are dependent focusing on the crucial phase of image acquisition [10]. In order to get the image obtained from a real world source to be converted into a digital image there are various devices available to capture the image [11] such as a digital camera and various operations like processing and compression can be carried out. The image captured, compressed, stored and displayed is generally known as Image acquisition

3.5 Preprocessing

Prior to image processing, preprocessing is employed to eliminate distortions [7] and other unwanted components, as well as to extract the correct part of the image. It entails a variety of processes and techniques performed to a picture to improve its quality, decrease noise, and get it ready for more processing or analysis. Various boundary detection algorithms as in [25–27] includes the removal of unnecessary elements and image compression.

3.6 Segmentation

We have used ResNet-50 (Residual Networks) which acts as a backbone for object detection, image segmentation and many computer vision applications. The ResNet network was developed by four researchers, Shaquing Ren, Kaiming He, Xiangyu Zunzhou and Jian Sun. It was the inaugural winner of the 2015 ImageNet challenge, achieving a 3.57% error rate. ResNet was developed to address the issue of vanishing gradients in deep neural networks. was also handled by the Resnet which is a very common problem raised by the number of increasing layers.

Neural networks have become more complex recently and the development of Deep Learning has seen the emergence of networks that span a range of layers, from a few to more than a hundred, such as VGG16. The primary benefit of a very deep network is its capacity to represent highly complex functions.

For example, an image can be taught characteristics at a variety of layers of abstraction, from edges at the lower layers to more complex features [9] at the higher layers.

Using a deeper network, on the other hand has some issues. The lack of gradients in large neural networks is a major obstacle to successful training. This is due to the slow rate of gradient descent in these networks, as the gradient signal is often reduced to zero in a short amount of time. To illustrate this more precisely, when backpropagating from one layer to the other, the weight matrix is multiplied at each step, meaning that the gradient may quickly decrease to zero, impeding the training process.

3.7 Identity Block

In ResNets, the identity block is the default block that is used when the input activation is the same size as the output activation .

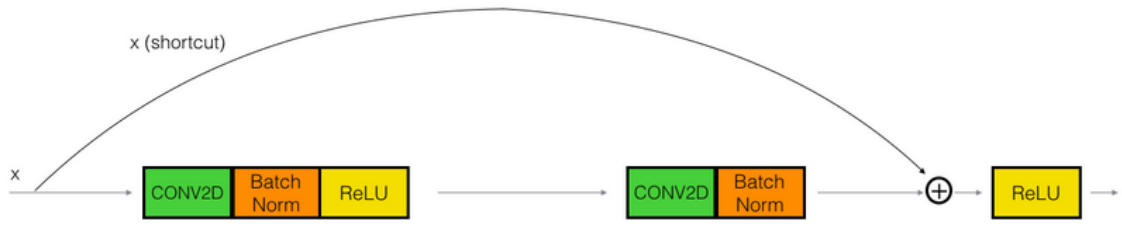

Figure 2: ResNet Identity block

3.8 Convolutional Block

This type of block can be used when the output dimension does not match the input dimension. The only difference between this type of block and an identity block is that the shortcut path contains CONV2D layers.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

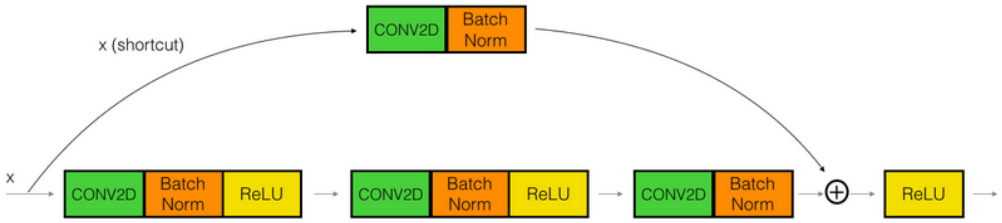

Figure 3: ResNet Convolutional block

3.9 ResNet-50

ResNet-50 has five layers, each with its own Convolution block and Identity block. Convolution blocks have three layers of convolution, while Identity blocks have three layers each. The model has over 23,000,00 trainable parameters

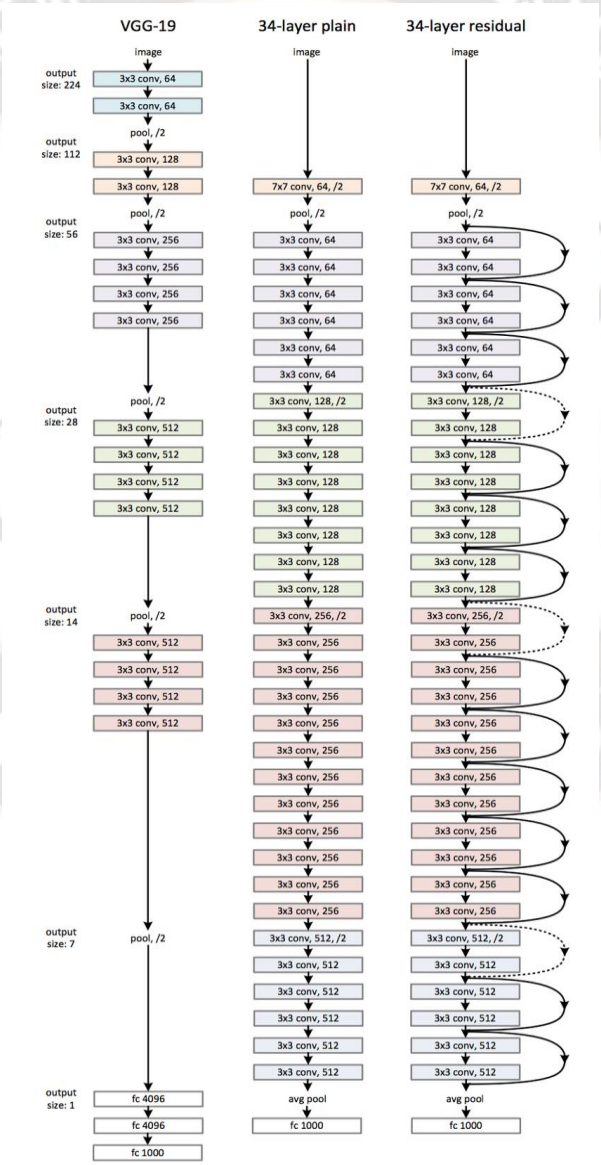

Figure 4: ImageNet network architecture. Left: the VGG-19 model (19.6 billion FLOPs) as a reference. Mid- dle:34 parameter layers plain network (3.6 billion FLOPs). Right: 34 parameter layers residual network (3.6 billion FLOPs). Other variants and detailing is presented in Table 1[29]

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Table 1: ImageNet architectures are characterized by their building blocks, which are indicated by brackets indicating the number of blocks that have been stacked. Conv3.1, 4.1 and 5.1 are used while downsampling with a sampling stride of 2 [29]

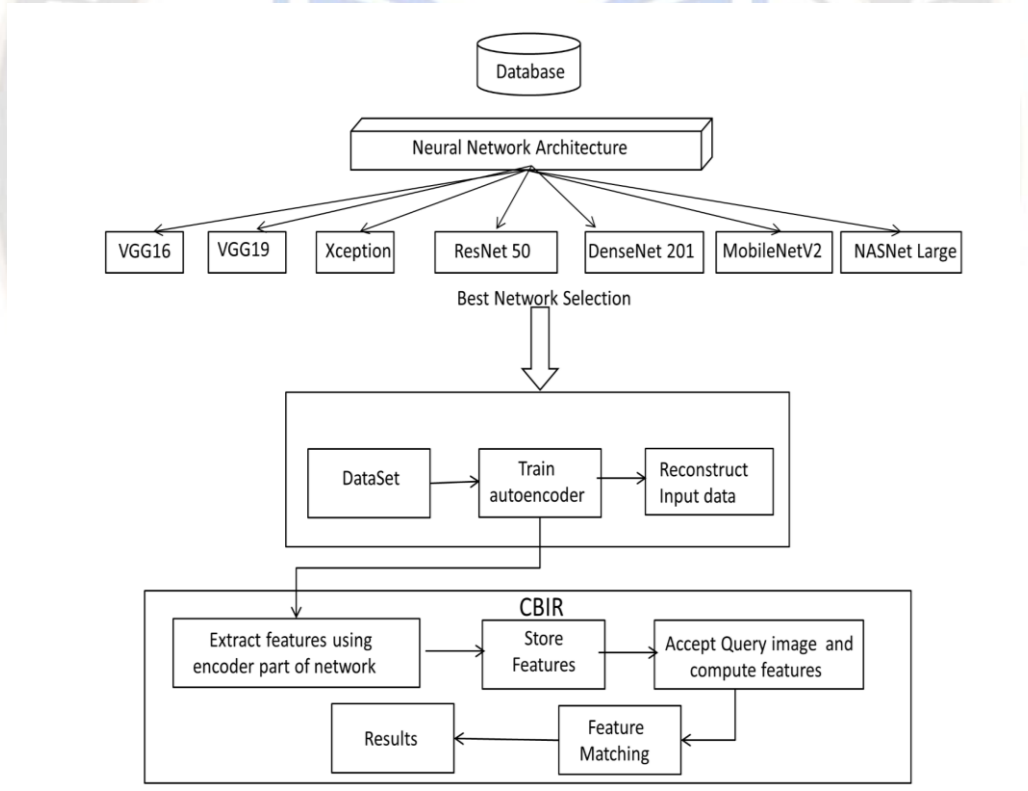

Figure 5: Proposed architecture for CBIR system using best network selection

This paper employed a feature extraction approach using only feature extraction layers. The most suitable feature extraction ImageNet dataset, with four parameters to be taken into account: This text examines the average values of four structure was compared using Table 2. The experiment was performed on an
parameters in an ImageNet for a variety of models. Loss and accuracy parameters for training and testing used to determine

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

the performance of the network. The network was trained to have at least 100 iterations, and 265 incremental steps were taken per epoch. Here, the average of the parameters for each model are presented

Sr. No	Model name	Train accuracy	Train loss	Test accuracy	Test loss
1	Zisserman et al VGG 16 (2014)	80.95%	08.83%	61.32%	19.47%
2	Zisserman et al VGG 16 (2014)	80.28%	09.42%	57.76%	19.52%
3	ResNet 50 He et al (2015)	80.94%	08.73%	54.34%	23.57%
4	Xception Chollet (2017)	76.33%	12.71%	53.21%	26.15%
5	Howard et al. MobileNetv2 (2017)	78.12%	9.00%	51.90%	25.51%
6	Huang et al. DenseNet201 (2017)	81.80%	09.14%	58.64%	19.30%
7	NASNetLarge Zoph et al. (2018)	76.74%	12.33%	44.59%	20.18%

Table 2: Comparison of average computed values of four parameters on ImageNet dataset

Sr. No.	Author	Year	Method/Methodology	Comments
1.	Sabry et al	2023	Maximized Convolutional Autoencoder Models for Large Datasets using InfoGAN and ViT.	Obtained better results for real facial images with huge datasets
2.	Chugh et al	2022	CDH(Color difference Histogram) was used for retrieval of images	The Euclidean distance method provided the highest performance among all distance methods due to its higher recall rate and accuracy than those of the other distance methods.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

3.	Arora et al	2022	Analyses were conducted on multiple deep learning structures using max-pool overlap pooling to fine-scale the networks, decision layers of the same size and type, with the same parameters across each network.	Obtained 99% precision for X- Ray based image retrieval for a given query image
4.	Kumar et al	2021	The segmentation process and the HaarDWT, as well as the lifting wavelet schemes, are utilized for the extraction of features in CBIR.	Better performance received over the previous Haar Wavelet schemes
5.	Yu et al	2021	Used partial differential equations that replicates how a person perceives images	Development and innovation can be accelerated and improved for E-Commerce platforms and online shopping should be promoted more effectively.
6.	Jaya et al	2020	Feature extraction, description, and matching techniques for images that have been developed during the previous 10 to 15 years using local features and low-level feature contents.	Global feature extraction techniques are well-suited for capturing images of natural landscapes, but they perform less well with images of structures and items that were created by humans.
7.	Pathak et al	2021	Extracting features for the CBIR model is proposed to use an enhanced version of darknet-53 known as group normalized-inception-darknet-53. This is known as gn-inception- darknet-53.	The suggested method outperformed the nineteen methods that were utilized for comparison, which included conventional and CNN methods for CBIR, for all of these datasets.
8.	Shukran et al	2021	CBIR uses a variety of criteria including color, texture, form, and integrated features (hybrid approaches). Precision, re-call, and response speed are the criteria.	Obtained better precision for combined feature technique
9.	Subhadeep et al	2020	Features produced from deep learning convolution network models that have already been trained to solve a significant image classification problem	Used image clustering technique to further reduce the retrieval time along with precision

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
***Article Received:** 30 July 2023 **Revised:** 18 September 2023 **Accepted:** 02 October 2023*

10.	Senthil et al	2019	The Histogram Search Algorithm will spread an image based on its color distribution.	Examined various techniques for extracting the pertinent low-level information and a number of interval measures to compare images.
11.	Thepade et al	2015	Haar Wavelet Transform (HAAR) and Canny Edge detector (CED) are capable of extracting Using the slope magnitude method, one may create form features with gradient operators like Sobel (Prewitt), Laplace (Laparte), and Frei (Chen).	This database contains 350 color images and offers superior performance compared to other algorithms that have been implemented using Canny technology.

Table 3: Summary of some of the recent trends in image retrieval and findings

3.10 Model Training for Image Classification

The following steps can be used to extract features from the ResNet model that has been pre-trained.

a. We use the imageNet classification dataset to train ResNet50.
b. PyTorch is the framework for downloading the pre-trained ResNet50 model.
c. The characteristics obtained via the fiinal fully coupled layer are utilized to train a multichannel SVM classification model.
d. The data loader is used for feeding the training and Testing datasets.
e. The Input data and model are utilized for the extraction of features.
f. The features have been visualized.
g. All training and testing images are added to the frontend, and each embedded image is saved.
h. Stored images are loaded into a trained resnet 50.
i. We load the model after that.
j. Standard scalers are utilized to scale the train loaders and The test loaders. Subsequently, the two datasets are connected independently.

Data Visualization:

A distance metric is used to determine how similar (or dissimilar) two database images (I) stored in the system are to each other. It is anticipated that this distance will appropriately gauge how different (or similar) the photos are from one another. Greater similarity is implied by a smaller computed distance. The degree of similarity [10] between two photographs may vary depending on how each user interprets them. Geometrically and probabilistically, consists of similarity measures mainly of two types. The former relies on the separation of feature vectors, while the latter is based on a probabilistic approach, even though previous research [3] suggests that probabilistic methods are effective. The two most popular Minkowski model used for comparison of query images and database images are L1 and L2 (L1-norm, L2-norm). Gaussian classifications are frequently employed to classify a pair as relevant or irrelevant, with higher likelihoods classified as relevant and lower likelihoods classified as irrelevant. They are computationally expensive, yet they perform substantially better than the geometric approaches [11]. L1 or L2 norms have been employed as the dissimilarity metric throughout this paper. The difference between the features in the database image P and the query picture S is called the Manhattan Distance or L1 norm which can be calculated as:

$$\boldsymbol{D(P,S) = \sum_{n=1}^{n} |xi, P - Xi, S|}$$

Euclidean distance between P and S, or L2 norm, is given by

$$\boldsymbol{D(P,S) = \sum_{n=1}^{n} (xi, P - Xi, S)}$$

The image's feature dimension is represented by n, with x representing the feature value of i, P representing the number of images, and x representing the number of features of the i-th image for both the query image and the database image's feature value.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

3.11 Performance Measure

Precision is the most frequently employed metric to assess the effectiveness of a Content Based Image Retrieval System. Precision is defined as:

$$\text{Precision} = \frac{\textit{Number of relevant images retrieved}}{\textit{Number of retrieved images}} \quad (3)$$

CBIR techniques typically return a pre-defined positive integer count of images, known as the system's scope. The precision value of each image within the database is then computed for an average of overall images within the database. Generally, the wider the scope is, the more relevant images are retrieved, leading to lower precision numbers.

$$TPR = \frac{TP}{Actual\ Positive} = \frac{TP}{TP + FN}$$
$$FNR = \frac{FN}{Actual\ Positive} = \frac{FN}{TP + FN}$$
$$TNR = \frac{TN}{Actual\ Negative} = \frac{TN}{TN + FP}$$
$$FPR = \frac{FP}{Actual\ Negative} = \frac{FP}{TN + FP} \quad (4)$$

IV. RESULTS AND DISCUSSION

The overall results obtained are presented here, Which includes the optimal architecture using deep learning; efficiency of retrieval in our CBIR system in comparison to given query images extracted across the datasets; the accuracy of category level retrieval of images in relation to datasets used ; and a proposed method's comparison accuracy against some other recent CBIR approaches.

4.1 Optimal Selection of Deep Learning Network architecture

An optimal deep learning architecture is selected by analyzing the network as outlined in Section 3.9. On the ImageNet dataset, precision average was calculated for each of them at an average scope value of 20. The Euclidean Distance metric (L2 norm) was utilized to determine the dissimilarity metric. According to the data presented in Figure 7, the average precision rating for each of them was 96.115%, indicating that ResNet 50 significantly outperformed the others. Consequently, for all subsequent trials, the chosen network design was ResNet50.

4.3 Image Retrieval of Sample Query Images

This query image has a precision of 1, and the top 5 results are obtained and presented in Figure 11, using the Corel dataset for scope 20, the architecture of ResNet 50, and a sample query image as shown in Fig. 11. The query images are classified into various categories and each result is relevant to the corresponding query image.

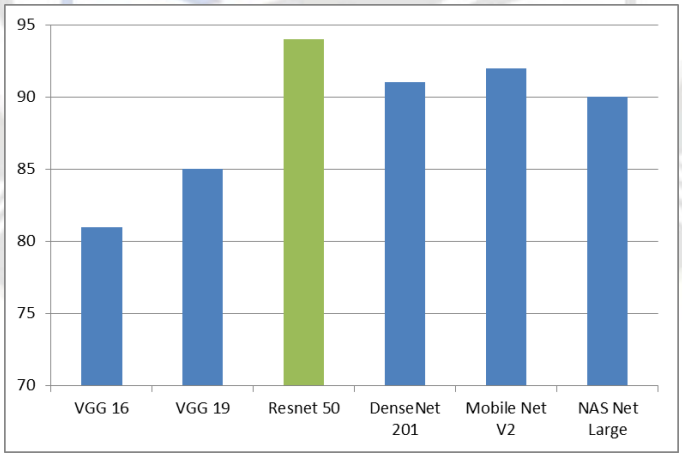

Figure 6: Corel Dataset compared on Deep Learning Network Architecture

Throughout the research, we draw on four distinct datasets, each of which has a unique combination of categories and image kinds. With 80% of the training photos and% of the test photos, the suggested models partition the datasets into training and test images. The first collection of data is Corel 1K [40], which consists of 1000 images split into ten groups.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

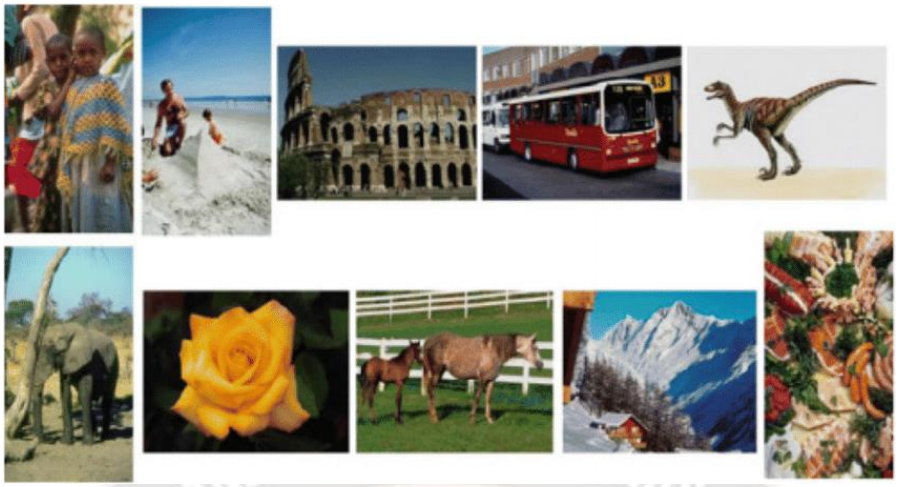

Figure 7: Sample images from Corel 1k [16]

The second dataset is CIFAR-10 [41]. There are ten categories and 60000 pictures in this public collection. The image is 32x32 pixels in size. Every class contains 6000 images. A training phase with 50,000 photos and a testing phase with 10,000 photos make up Cifar10.

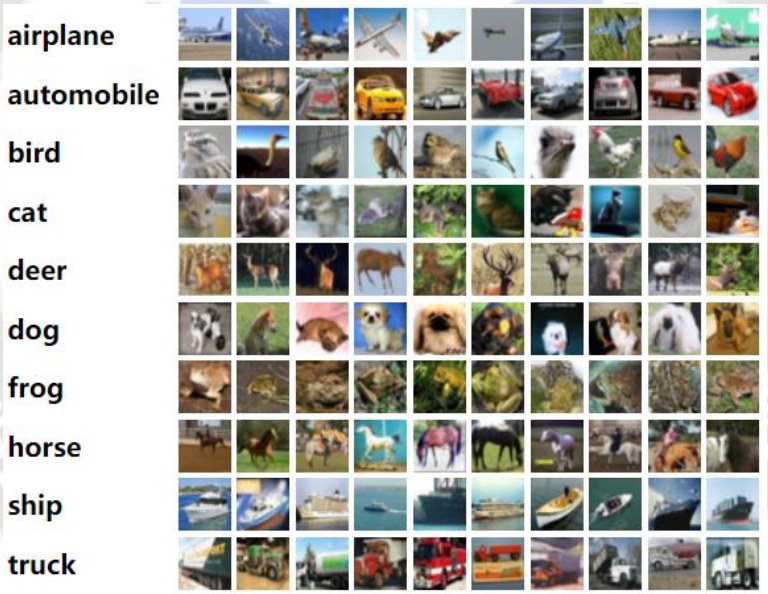

Figure 8: Sample images from Cifar 10 [17]

CIFAR 100 is a dataset that consists of 100 classes, each of which contains 600 photos. Within each class, 100 assessment photos are included, while 500 training images are included. The dataset is composed of 20 super classes, each of which is divided into 100 classes. Each image is labeled with either a fine or coarse label, indicating which class it belongs to and which superclass it belongs to.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

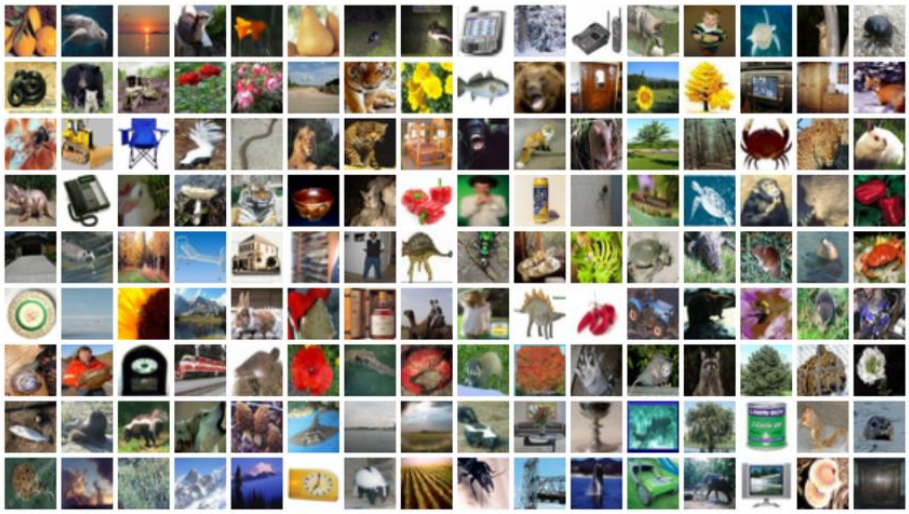

Figure 9: Sample images from Cifar 100 [17]

4.4 Category wise precision calculation

The measure of dissimilarity is calculated by Euclidean Distance to generate the Category-wise Average Precision (CAMD) on Corel (1k) as well as CIFAR (10) with a range of 20 for this subsection. The results of the study are presented below in Table 1 and Figure 16. The results demonstrate that the ResNet50 Model pre-trained in Corel (1K) retrieves all photos with absolute accuracy for the leaf, flower, and aeroplane categories. but performs comparably badly for the category of fruit, with a precision of 77.35%. This dataset's total average precision is 96.115%. The categories with the highest precision in CIFAR 10 are Frog (precision: 99.95%) and Automobile (precision: 99.20%), with an overall average precision of 97%. Table 4 and Figure 16 shows the results obtained.

Categories	Average Precision %
Flower	100
Fruit	77.35
Car	87.20
Animal	92.30
Aeroplane	100
Ship	82.55
Leaf	100

(a)

Categories	Average Precision %
Airplane	87
Automobile	99.20
Bird	92
Plane	82
Dog	92
Frog	99.95
Horse	94

(b)

Table 4: Average precision obtained for **(a)** Corel 1k and **(b)** CIFAR 10 Datasets

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

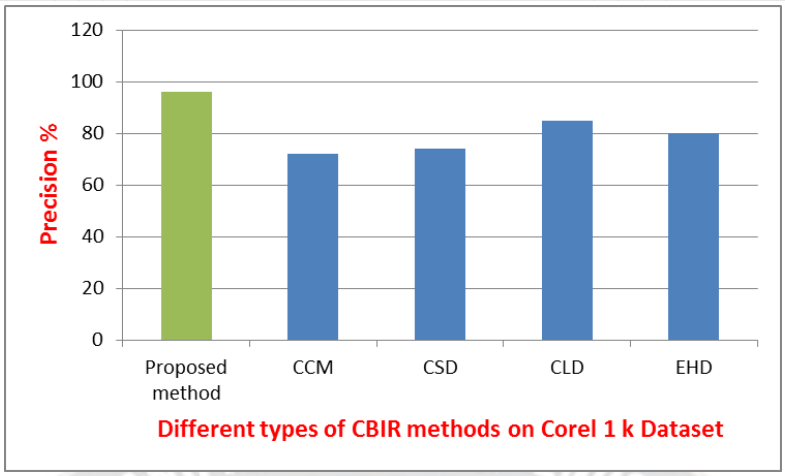

Figure 10: Comparison between the average accuracy of our method and the average accuracy of Bose et al [7] using Corel 1K Dataset for a scope of 20.

Figure 11: Top 5 image retrieval results for the Corel1K dataset

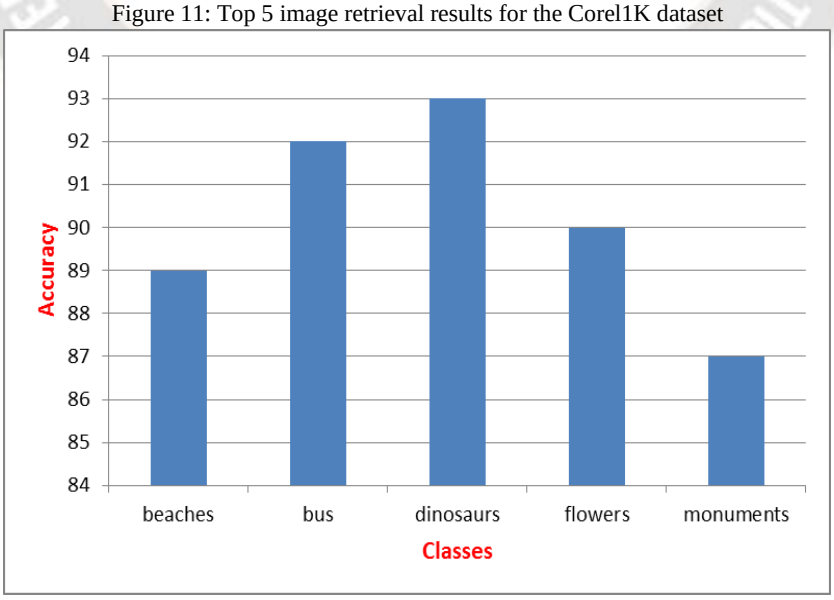

Figure 12: Category wise results obtained for Corel 1K dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

Figure 13: Top 5 image retrieval results for the Cifar-10 dataset

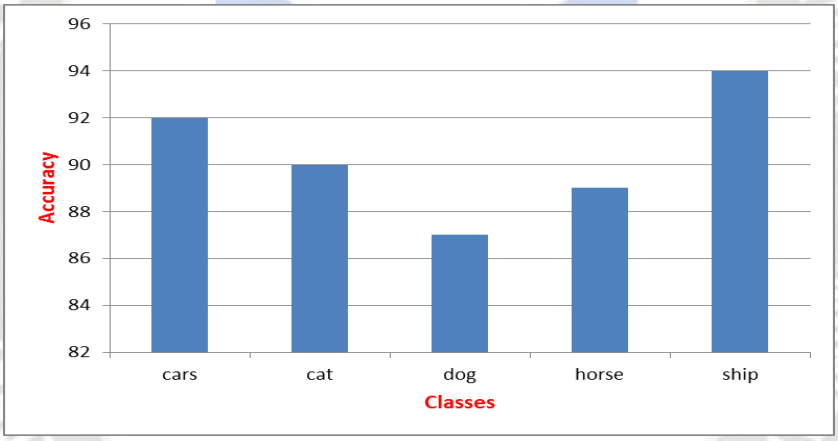

Figure 14: Accuracy obtained for various classes of CIFAR 10 dataset

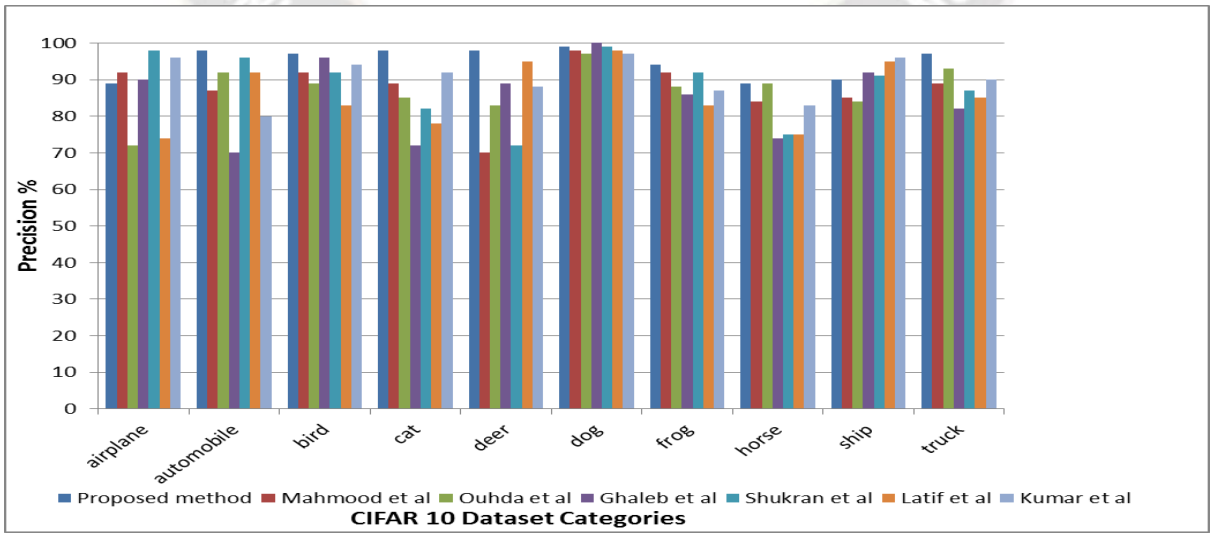

Figure 15: Comparing the results to other recently proposed algorithms

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

4.5 Comparison of results with other recently proposed algorithms

Gahaleb et al.[19] is the primary source of data for the Corel 1K Dataset. This paper provides an analysis of three proposal methodologies based on deep learning: CNN, Convolutional LSTM Fused and GRU. As this paper does not provide relevance for feedback, the accuracy of this paper compares without relevance feedback to our proposed method. Figure 15 demonstrates that our proposed approach outperforms all methods discussed in this paper [20]. In the past few years, a number of research papers have been conducted on Corel Dataset, which extracts various types of features and distances between them. This paper compares two types of methods published in these recent papers, Comparison of category wise precision (Fig. 16) and calculated Average Precision comparison (Table 5) to our proposed method. Outperforming algorithms and methods published in these papers, it is evident that the proposed method gives far better results. For instance, according to Lohite et al [25] computed Precision for scope of 50 rather than scope of 20 among various categories. Additionally, a comparative analysis was conducted using this algorithm (see Fig. 16), and it was observed that, except for the category African People, our proposed algorithm is more efficient for scope of 50.

In order to provide a comparative analysis for the CIFAR 10 dataset, we generate the comparison using two algorithmic methods as described in [7] and [29,30]. Calculated average precision of the two best methods is compared in Figure 20. We have discussed multiple CBIR techniques for the two methods in the paper thus we have selected the average precision for the proposed method. It is evident that the proposed method significantly outperformed the two other methods.

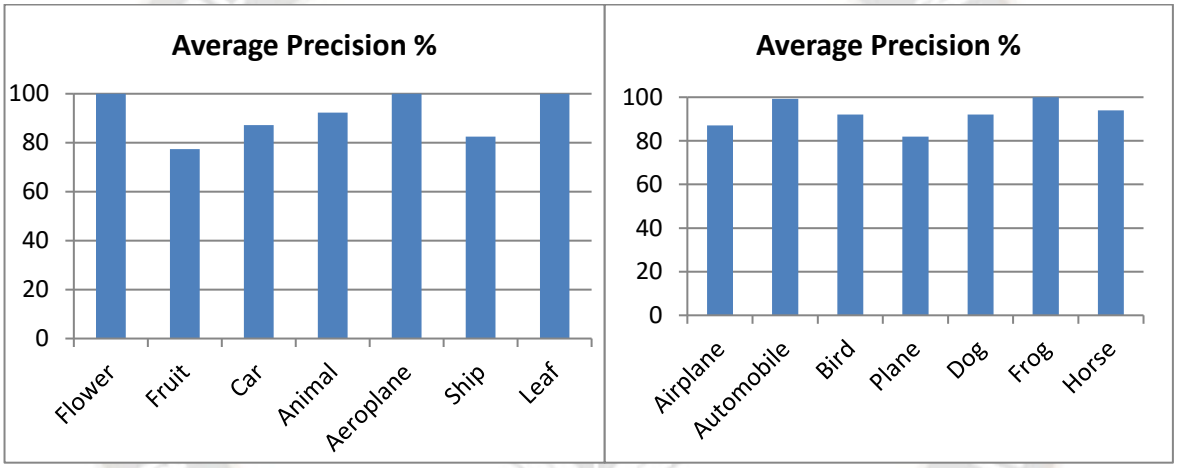

Figure 16: Category wise average precision for scope of 20 on (a) Corel 1k (b) CIFAR 10

Methods	Average Precision %
Proposed Method	97.12
Kumar et al	91.77
Chugh et al	96.00
Shukran et al	88.30
Ghaleb et al	93.30
Palai et al	94.50
Mahmood et al	95.70

Table 5: A comparison of the average accuracy of our proposed method in comparison to the methods used in recent papers for a scope of 20 on Corel 1k dataset

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

There are 1000 photos with dimensional features in Corel 1k and 10000 images with a total of 1536 dimensional elements. in CIFAR 10 (without PCA). We can see that our GPU computer, and to a lesser extent, our local machine, are likewise capable of instantly retrieving photographs from these datasets. The size and dimension of the database affect how quickly images can be retrieved. While utilizing the same architecture (ResNet50 in our case), the dimension is almost the same. Yet if we employ a dataset with millions of photographs in a collection, the retrieval time for each image inevitably increases. If it turns out that scanning through the entire database for pertinent photographs is taking a long time, It is possible to extract 20 images from a database containing millions or billions of images by using a random sample of 10,000 to 20,000.

V. CONCLUSION

This research demonstrates that, compared to features created using conventional techniques, such as CCM, wavelet, etc., employing features of a pre-trained deep learning network architecture yields greater precision results. However, adding user comments, also known as relevance feedback, can enhance this result for a particular dataset. After each retrieval, users provide input on which results are 15 relevant to the query photos and which are not. This response is known as "relevance feedback." The CBIR system will begin to learn from this feedback and eventually improve the outcome. The fact that deep learning-derived features are not rotation-invariant is one of its limitations. As a result, if we attempt to find related images using the same query image but different orientation angles, the retrieval outcomes will vary greatly each time. Making a rotation-invariant CBIR system could be the next stage in the development of this method.

REFERENCES

[1] Sabry, E. S., Elagooz, S. S., El-Samie, F. E. A., El-Shafai, W., El-Bahnasawy, N. A., El-Banby, G. M Ramadan, R. A. (2023). Image Retrieval Using Convolutional Autoencoder, InfoGAN, and Vision Transformer Unsupervised Models. IEEE Access, 11, 20445–20477. https://doi.org/10.1109/ACCESS.2023.3241858

[2] Arora, N., Kakde, A., & Sharma, S. C. (2022). An optimal approach for content-based image retrieval using deep learning on COVID-19 and pneumonia X-ray Images. International Journal of System Assurance Engineering and Management. https://doi.org/10.1007/s13198-022-01846-4

[3] Eom, M., & Kim, B. I. (2020, December 1). The traffic signal control problem for intersections: a review. European Transport Research Review. Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1186/s12544-020-00440-8

[4] Sundararajan, S. K., Shankaragomati, B., & Priya, D. S. (2019). A performance perspective: Content based image retrieval system. International Journal of Recent Technology and Engineering, 7(6), 1547–1555.

[5] Maji, S., & Bose, S. (2021). CBIR Using Features Derived by Deep Learning. ACM/IMS Transactions on Data Science, 2(3), 1–24. https://doi.org/10.1145/3470568

[6] He, Kaiming & Zhang, Xiangyu & Ren, Shaoqing & Sun, Jian. (2016). Deep Residual Learning for Image Recognition. 770-778. 10.1109/CVPR.2016.90.

[7] Shukran, M.A.M., Abdullah, M.N. and Yunus, M.S.F.M. (2021) New Approach on the Techniques of Con- tent-Based Image Retrieval (CBIR) Using Color, Texture and Shape Features, Journal of Materials Science and Chemical Engineering,9,51-57, https://doi.org/10.4236/msce.2021.91005

[8] Zhao, Y., Wang, J., & Qi, Q. (2018). MindCamera: Interactive Image Retrieval and Synthesis. In 2018 IEEE 13th Image, Video, and Multidimensional Signal Processing Workshop, IVMSP 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/IVMSPW.2018.8448722

[9] Pathak, D., & Raju, U. S. N. (2021). Content-based image retrieval using Group Normalized-Inception-Darknet-53. International Journal of Multimedia Information Retrieval, 10(3), 155–170. https://doi.org/10.1007/s13735-021-00215-4

[10] Dewan, J. H., & Thepade, S. D. (2020). Image Retrieval Using Low Level and Local Features Contents: A Comprehensive Review. Applied Computational Intelligence and Soft Computing. Hindawi Limited. https://doi.org/10.1155/2020/8851931

[11] Chugh, H., Gupta, S., Garg, M., Gupta, D., Juneja, S., Turabieh, H., … Kiros Bitsue, Z. (2022). Image Retrieval Using Different Distance Methods and Color Difference Histogram Descriptor for Human Healthcare. Journal of Healthcare Engineering, 2022. https://doi.org/10.1155/2022/9523009

[12] Yu, J. (2021). Multifeatured Image Retrieval Techniques Based on Partial Differential Equations for Online Shopping. Advances in Mathematical Physics, 2021. https://doi.org/10.1155/2021/2834873

[13] Kumar, S., Jain, A., Kumar Agarwal, A., Rani, S., & Ghimire, A. (2021). Object-Based Image Retrieval Using the U-Net-Based Neural Network. Computational Intelligence and Neuroscience, 2021. https://doi.org/10.1155/2021/4395646

[14] Thepade, S. D., & Shinde, Y. D. (2015). Improvisation of content based image retrieval using color edge detection with various gradient filters and slope magnitude method. In Proceedings - 1st International Conference on Computing, Communication, Control and Automation, ICCUBEA 2015 (pp. 625–628). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICCUBEA.2015.128

[15] Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I. Khalil, T. (2019). Content-based image retrieval and feature extraction: A comprehensive review. Mathematical Problems in Engineering, 2019. https://doi.org/10.1155/2019/9658350

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
DOI: https://doi.org/10.17762/ijritcc.v11i10.8478
Article Received: *30 July 2023* ***Revised:*** *18 September 2023* ***Accepted:*** *02 October 2023*

[16] Taheri, Fatemeh & Rahbar, Kambiz & Salimi, Pedram. (2022). Effective features in content-based image retrieval from a combination of low-level features and deep Boltzmann machine. Multimedia Tools and Applications. 10.1007/s11042-022-13670-w.

[17] https://www.cs.toronto.edu/~kriz/cifar.html

[18] Mohamed, O., El Asnaoui, K., Mohammed, O., & Brahim, A. (2019). Content-Based Image Retrieval Using Convolutional Neural Networks. In Advances in Intelligent Systems and Computing (Vol. 756, pp. 463–476). Springer Verlag. https://doi.org/10.1007/978-3-319-91337-7_41

[19] Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2022). Image Retrieval Based on Deep Learning. Journal of System and Management Sciences, 12(2), 483–502. https://doi.org/10.33168/JSMS.2022.0226

[20] Mehmood, Z., Mahmood, T., & Javid, M. A. (2018). Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine. Applied Intelligence, 48(1), 166–181. https://doi.org/10.1007/s10489-017-0957-5

[21] Rashno, A., & Sadri, S. (2017). Content-based image retrieval with color and texture features in neutrosophic domain. In 3rd International Conference on Pattern Analysis and Image Analysis, IPRIA 2017 (pp. 50–55). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/PRIA.2017.7983063

[22] Rupapara, V., Narra, M., Gonda, N. K., Thipparthy, K., & Gandhi, S. (2020). Auto-Encoders for Content-based Image Retrieval with its Implementation Using Handwritten Dataset (pp. 289–294). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/icces48766.2020.9138007

[23] Palai, C., Jena, P. K., Pattanaik, S. R., Panigrahi, T., & Mishra, T. K. (2023). Content-based Image Retrieval using Encoder based RGB and Texture Feature Fusion. International Journal of Advanced Computer Science and Applications, 14(3), 245–254. https://doi.org/10.14569/IJACSA.2023.0140328

[24] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: arXiv e-prints, arXiv:1512.03385 (Dec. 2015), arXiv:1512.03385. arXiv: 1512.03385 [cs.CV].

[25] Mr. Yogen Mahesh Lohite and Prof. Sushant J. Pawar. "A novel method for Content Based Image retrieval using Local features and SVM classifier". In: International Research Journal of Engineering and Technology (IRJET) 04.07 (2017).

[26] Wayne Niblack et al. "The QBIC Project: Querying Images by Content, Using Color, Texture, and Shape." In: vol. 1908. Jan. 1993, pp. 173–187.

[27] Zahid Mehmood, Toqeer Mahmood, and Muhammad Arshad Javid. "Content-based Image Retrieval and Semantic Automatic Image Annotation Based on theWeighted Average of Triangular Histograms Using Support Vector Machine". In: Applied Intelligence 48.1 (Jan. 2018), pp. 166–181. ISSN: 0924-669X. DOI: 10.1007/s10489- 017-0957-5. URL: https://doi.org/10.1007/s10489-017-0957-5

[28] Smarajit Bose et al. "Improved Content-Based Image Retrieval via Discriminant Analysis". In: International Journal of Machine Learning and Computing 7 (June 2017), pp. 44–48. DOI: 10.18178/ijmlc.2017.7.3.618.

[29] Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun ,Microsoft Research, arXiv:1512.03385

[30] Kazmi, S., Singh, M., & Pal, S. (2023). Image Retrieval Performance Tuning Using Optimization Algorithms. International Journal of Experimental Research and Review, 33, 8-17. https://doi.org/10.52756/ijerr.2023.v33spl.002