IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

IMAGE RETRIEVAL SYSTEM USING DEEP LEARNING

[1]Smit Babariya, [2]Dr. Udesang K. Jaliya, [3]Dr. Mahasweta J. Joshi
[1]PG Scholar, [2] Assistant Professor, [3] Assistant Professor
[1] Computer Engineering,
[1] BVM Engineering College, Vallabh Vidyanagar, Anand- 388120, Gujarat, India

Abstract: Locating large picture collections requires the use of content-based image retrieval systems. Convolutional Neural Networks (CNNs) are deep learning algorithms which have upgrade the CBIR with their immense picture feature extraction capabilities. The efficiency of CNNs, ResNet-50, and VGG16 three prominent CNN architectures—for image retrieval is studied in this work. This is investigating a transfer learning-based deep learning framework for CBIR. In this study pre trained models such as CNN, ResNet-50 and visual geometry group (VGG-16) are used to obtain properties from image. These models were trained using large Cifar-10 image datasets. These attributes allow for efficient retrieval based on similarity by capturing the visual properties and semantic content of the images. The study evaluates how effectively these CNN architectures perform in terms of retrieval accuracy. A versatile framework for modifying feature extraction layers is provided by standard CNNs. With its residual connections, ResNet-50 enables deeper structures and better learning. VGG-16's depth can cause overfitting on smaller datasets. The results of this study can help in the development of reliable and effective Content based image retrieval systems. For researchers and practitioners looking to use deep learning for image retrieval applications, the comparison offers valuable insight.

Keywords: CBIR, CNN, Pretrained RESNET-50 model, Pretrained VGG16 model, Cifar-10 dataset.

I. INTRODUCTION

Effective image retrieval systems are becoming more necessary in this era of abundant information, since digital photographs are widely used in fields like social media. Conventional image retrieval techniques frequently fail to grasp the subtle nuances present in photos because they depend too much on hand-crafted features and shallow learning models. A breakthrough era for image retrieval systems is being heralded by the emergence of deep learning, especially through Convolutional Neural Networks (CNNs), which promise improved accuracy and durability through automated feature learning [3].

This conceptual shift deviates from traditional methods because deep learning enables systems to extract precise visual representations on their own. Deep learning satisfies the growing need for more complex and versatile image retrieval solutions by automating feature learning. Essentially, this revolutionary stage not only tackles the shortcomings of conventional approaches but also synchronizes image retrieval systems with the changing requirements of a dynamic and visually stimulating digital environment.

The size of picture collections highlights how crucial it is to develop efficient methods for retrieving relevant photos in response to user inquiries. Conventional techniques rely on hand-crafted attributes that might not fully capture the full visual information which is present in images. The limitations of these traditional approaches become increasingly apparent as the amount of picture data increases rapidly, underscoring the need for sophisticated and adaptable solutions.

It is obvious that the application of hand-crafted features has limited capacity for properly capturing the depth of visual data given the growing volumes of picture information. The desire for advanced and flexible approaches is a reflection of the need to keep up with the growing diversity and complexity of image collections. New approaches, particularly ones based on machine learning and deep learning, must be investigated and implemented in order to address the increasing problems with image retrieval in our more picture-centric digital world feature extraction method: The basis of deep learning-based picture retrieval methods is convolutional neural network ability to extract an extensive variety of intricate features from images. Unlike previous approaches that depend on features that are human-engineered, CNN automatically capture low-level data and high-level definitions [2]. Analyzing the context and content of photos involves the ability to extract features from image.

Training Data: Huge volumes of training data are ideal for deep learning models. CNNs are trained on large datasets of tagged pictures so they may acquire discriminative representations. In order to maximize its capacity to map pictures into the intended embedding space, the network optimizes its parameters during training to distinguish between various images.

Current image retrieval systems often struggle with semantic gaps and lack the flexibility to understand nuanced user intent. This project aims to develop a deep learning-powered image retrieval system that goes beyond basic keyword matching, utilizing advanced attention mechanisms and context understanding to accurately retrieve images that truly reflect user query and intent, bridging the semantic gap and unlocking the full potential of image search [5].

Large image collections can be useful tools for deep learning model training, enabling the models to organize new photos with accuracy. The capabilities of deep learning models can be used by a different technique, including similarity search and Content-Based Image Retrieval. One approach uses ResNet-50 in conjunction with several Convolutional Neural Networks to do feature classification and extraction [6]. These models are subsequently employed to find relevant photos in response to search requests.

2.Content based Image Retrieval proposed Models

The two steps of the intelligent CBIR models are the training phase and the retrieve phase. A CNNs style deep learning algorithm was used in its training phase to identify crucial features like color, edges, and materials. As a result, a feature vector was created and used to group photos into several groups based to the attributes that could be obtained. The second stage is called retrieving. Once relevant images for the query picture have been retrieved from the database, the retrieval phase employs evaluation metrics to evaluate the model's performance. The operational structure of the CBIR intelligent method is shown in Fig. 1. Testing as well as training are the two main phases of the deep learning process.

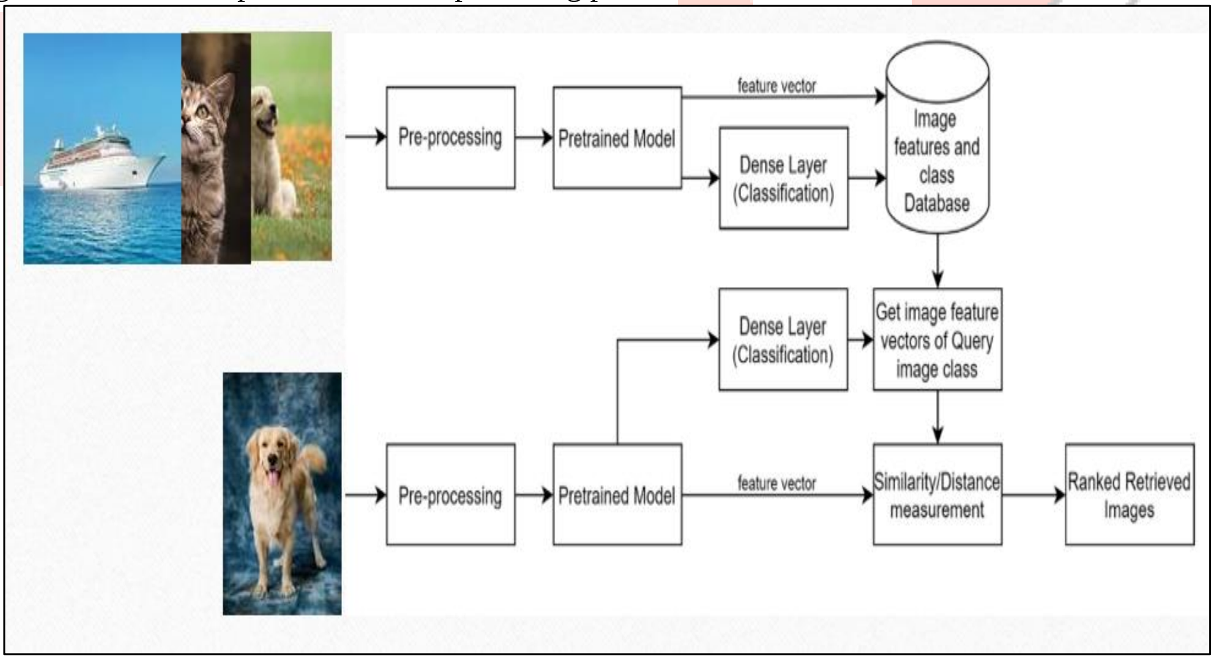

figure 1 block diagram of intelligent CBIR approach [4]

The CBIR strategy will change as a result of the usage of deep learning [1]. Feature extraction, image classification, and image classification are the three stages of the intelligent CBIR technique that classify the images. In the test phase, the category for the query image is forecasted. The forecast's validity is then assessed, and the average accuracy of the model is determined. At last, the test phase is used to obtain visual data from the dataset according to the classification class and use a few previously unseen images as a query image to assess the model's performance. The number of relevant images that were recovered is then calculated using the picture retrieval metrics. The performance of CBIR may be affected by deep learning methods like CNN, RESNET-50, and VGG16, which have demonstrated good classification accuracy in images. The objective of this work is to determine the best deep learning approach for resolving CBIR problems.

3. Deep Learning Proposed Models

3.1 Classification

The initial stage, which is called the classification step, uses supervision to teach the model how to classify images that are randomized. In this step, picture characteristics are determined and stored into a dataset index. The classification model is divided two phases: first, image attributes are collected, and then those features are used to categorize images. Content based image retrieval systems usually only include the convolutional neural network-based feature extraction phase. However, the classification model can still be utilized to categorize query photos by utilizing these attributes that are taken out of the CNN's deep layers. Using the expected class of the query image may facilitate the retrieval of similar images. Deep CNNs usually consists of multiple layers. The convolution stage includes layers to the input picture in order to obtain information from it. Every node in a totally linked layer is linked to every other layer's node.

3.2 Convolution neural network (CNN)

CNN's are a fundamental component of modern deep learning techniques for identifying images. Convolutional, pooling, and fully-connected layers make form the efficient framework used by convolutional neural network; this design is modeled by the structure of the cortex of the brain.

CNNs' crucial component, the convolutional layers, are in control of extracting features from pictures. They use filters that move over the picture to identify lines, edges, and other subtle details found in the first few layers. These filters gain the ability to recognize increasingly complicated elements as the network develops and combine them to create higher-level representations. Pooling Layers Utilizes down sampling techniques, pooling layers minimize the dimensionality of the data while maintaining crucial details. This aids in keeping the model's complexity under control and prevents against overfitting.

3.3 RESNET-50

A deep residual network architecture titled ResNet-50 was unveiled by Microsoft Research in 2015. It improves the essential CNNs architecture by solving the vanishing gradient problem, a common problem in training very deep networks.

The phenomenon called as the "vanishing gradient problem" occurs when gradients propagate backwards through the network during training and end up being unusually small or significant. The network finds it difficult to learn from deeper levels as a result. ResNet50 presents residual learning as a way to address this problem. The residual block is ResNet50's core building block. These blocks contain a lot convolutional layers that process the input data. A shortcut joins the original input to the convolutional layers' output directly. ResNet50 can efficiently train very deep models because to this shortcut, which guarantees that the gradients can flow back through the network without vanishing or bursting [6].

3.4 VGG16

Oxford University's Visual Geometry Group developed the popular deep learning image recognition method VGG16. The main advantages of its deep convolutional technique are its simplicity and effectiveness. With 16 convolutional layers built one after the other and using small 3x3 filters, the model is remarkable. This allows the receptive field to gradually expand, allowing VGG16 to capture more complicated details in the image. VGG16 creates a rich hierarchy of feature representations by stacking these layers, which eventually improves the accuracy of image categorization [4]. VGG16 has limits even if it performs admirably on benchmarks like the ImageNet challenge. Although useful for feature learning, its depth comes at a high computational cost. Utilizing and training VGG16 takes a significant number of resources.

In the field of deep learning, VGG16 is still a helpful model in spite of these limitations. According to ResNet-50 success in image recognition, improved architectures were able to solve VGG16 training problems. Deep convolutional networks are effective, as shown by VGG16, which maintains to push limits of computer vision research.

4. Data set

For use in image retrieval networks, many datasets are available. The picture retrieval system is available in multiple implementations, each using a different dataset. The Place 365, Cifar-100, and Cifar-10 random image collections are available.

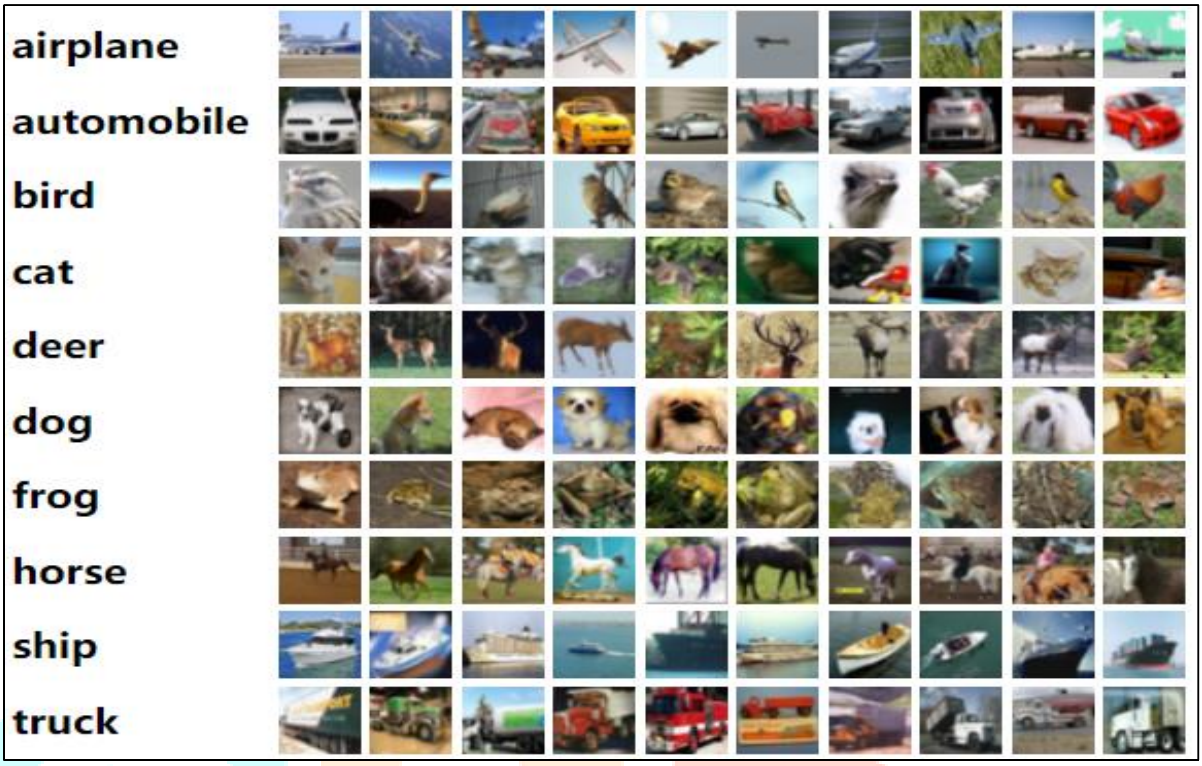

Figure 2 Sample for Cifar-10 data set [7]

9000 photos in 10 classes make up the Cifar-10 dataset. There are 90000 photos in the dataset, 9000 of which are for testing and 90000 for training.

5. Results

5.1 Results of classification model

The results with 29 epochs, the accuracy of the model is 0.7616. The dataset which is used to train the model is Cifer10.

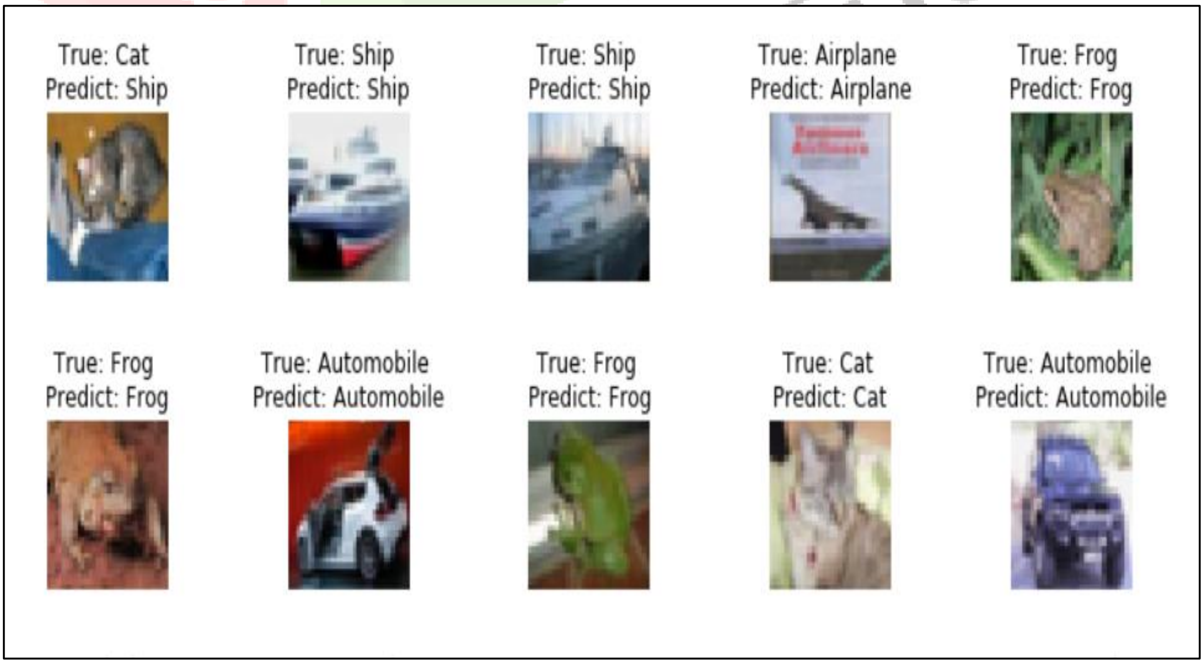

Figure 3 Results with 29 epochs

The results with 47 epochs, the accuracy of the model is 0.7861. The dataset which is used to train the model is Cifer10.

Figure 4 Results with 47 epoch

Figure 5 and Figure 6 shows model accuracy and loss accuracy for classification models.

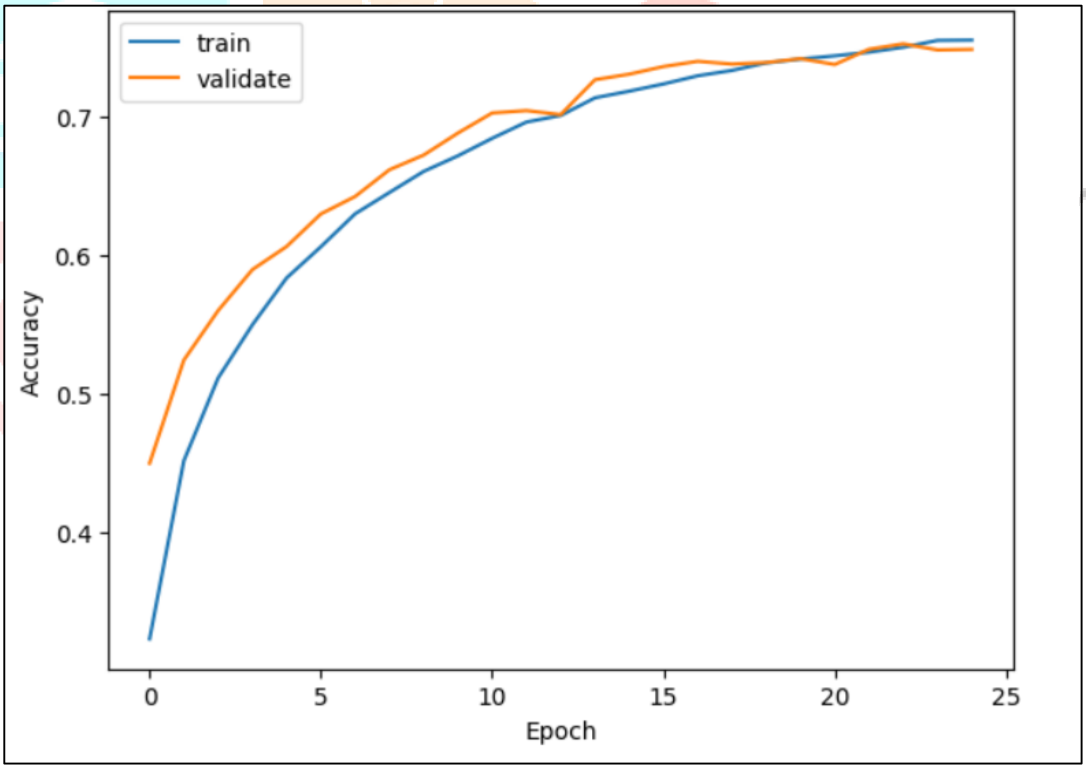

Figure 5 Model accuracy

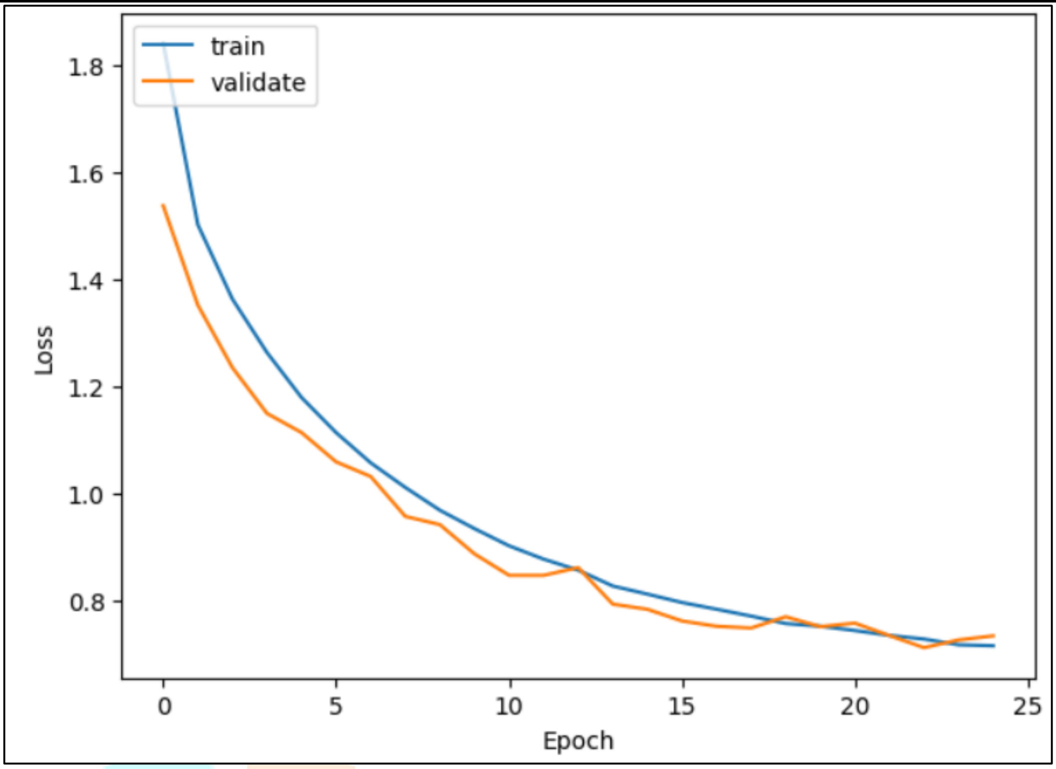

Figure 6 Loss Acuuracy

RESULT OF RESNET-50 MODEL:

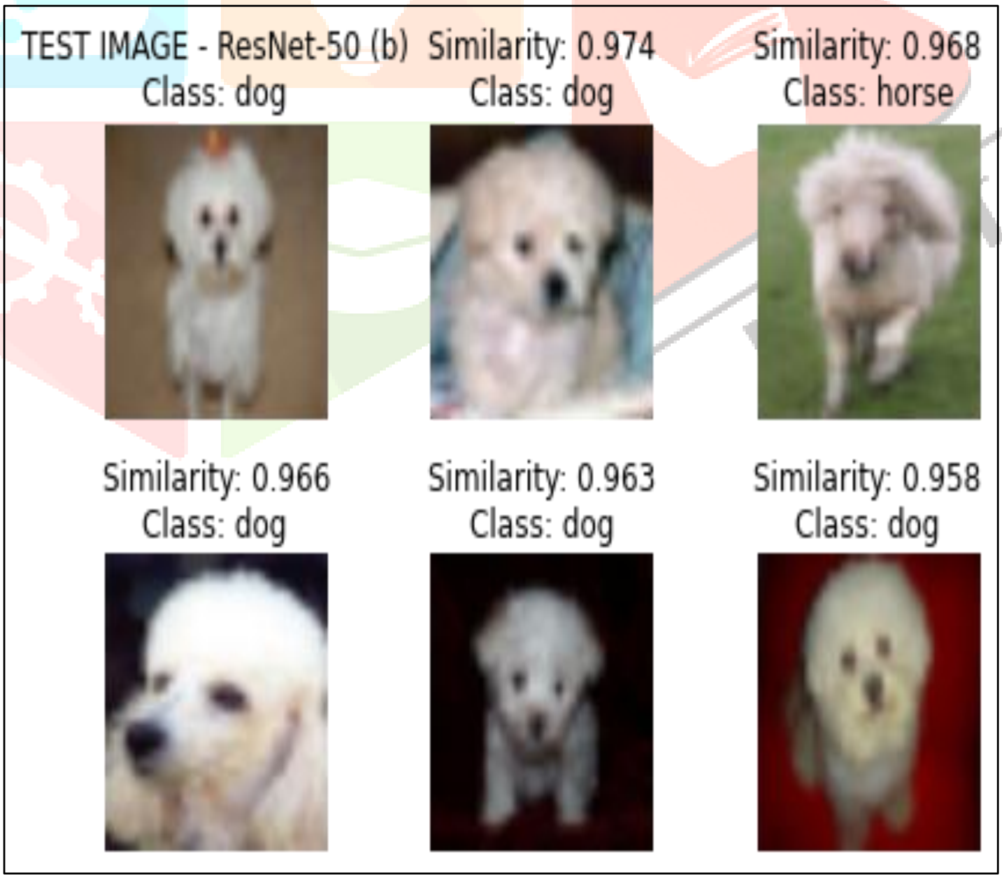

Figure 7 Result of ResNet-50 Model

Figure 8 and Figure 9 shows the confusion matrix and accuracy of RESNET-50 model respectively.

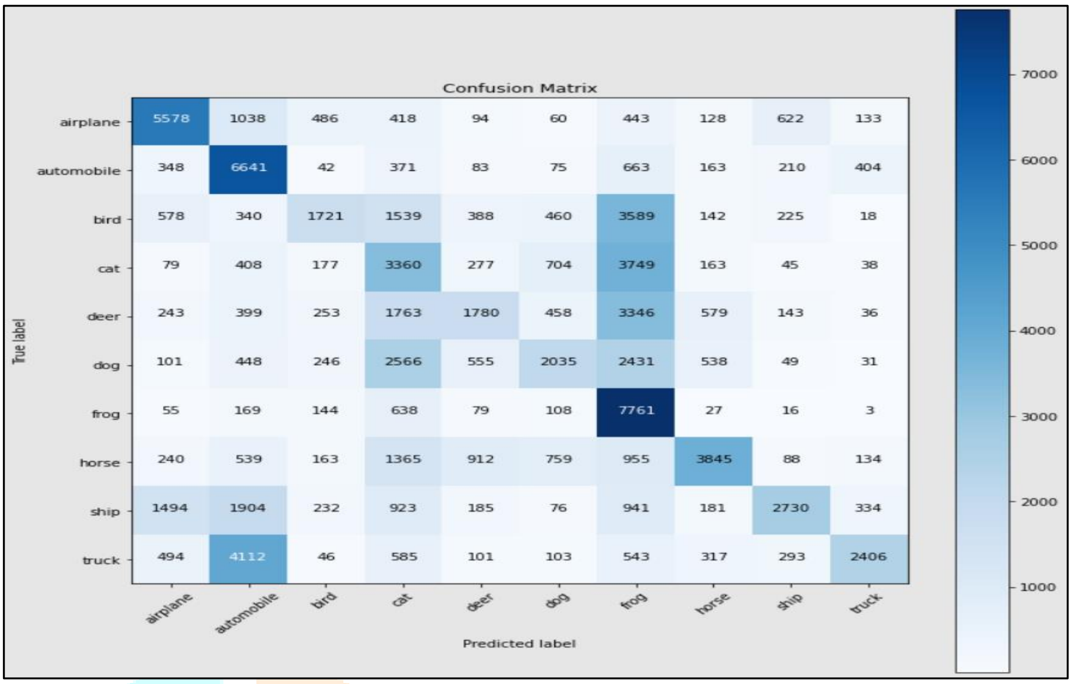

Figure 8 RESENET-50 Confusion Matrix

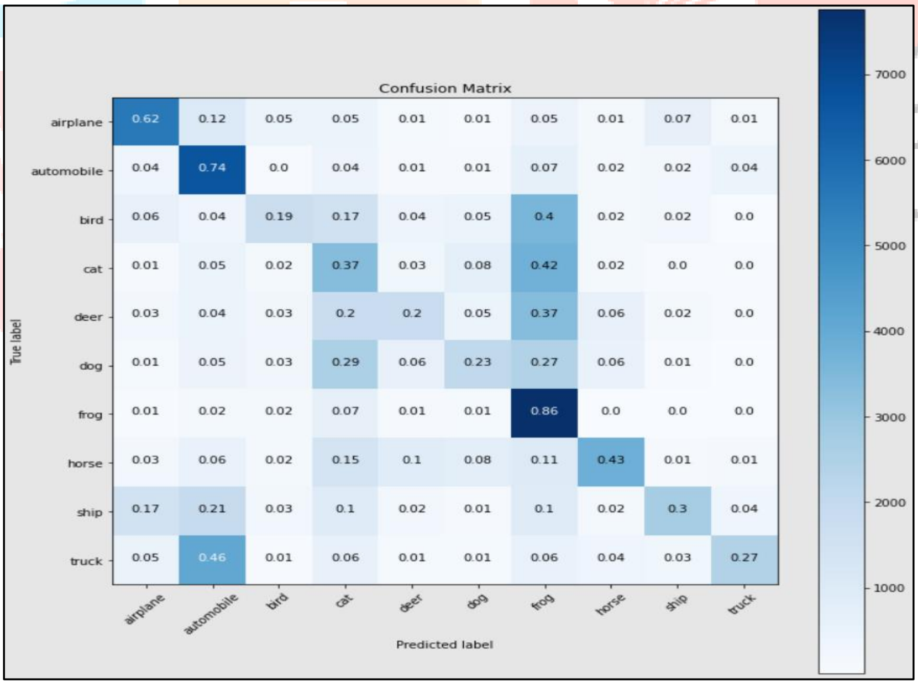

Figure 9 Accuracy of RESNET-50 Model

RESULT OF VGG-16 MODEL:

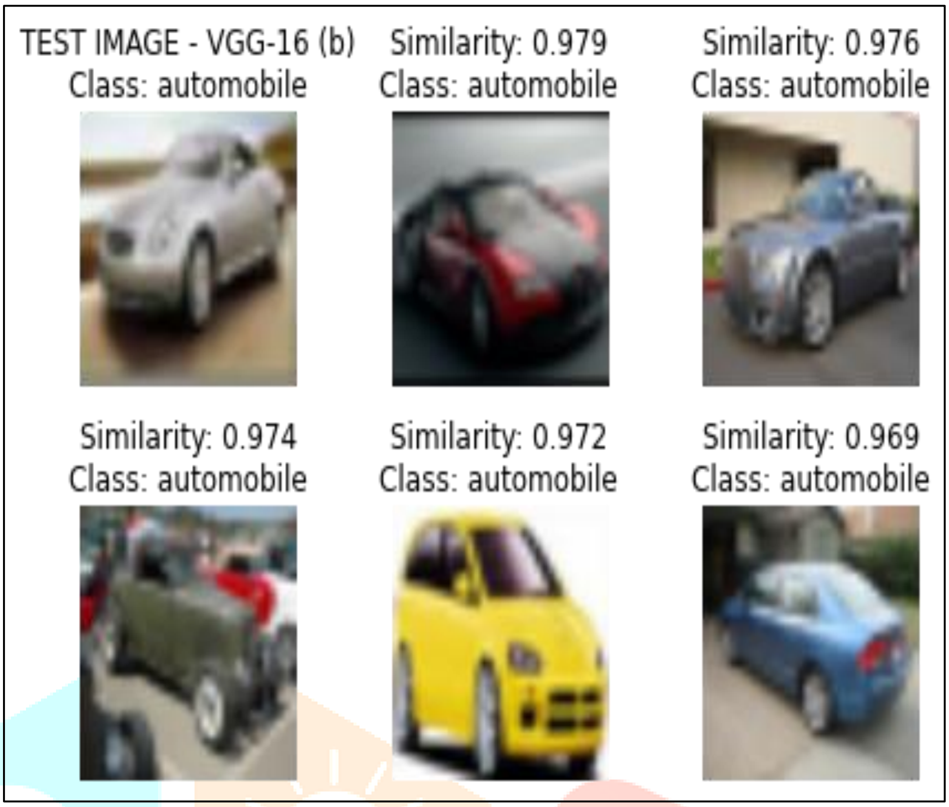

Figure 10 Result of VGG16 Model

Figure 11 and Figure 12 illustrate the confusion matrix and accuracy of VGG16 models respectively.

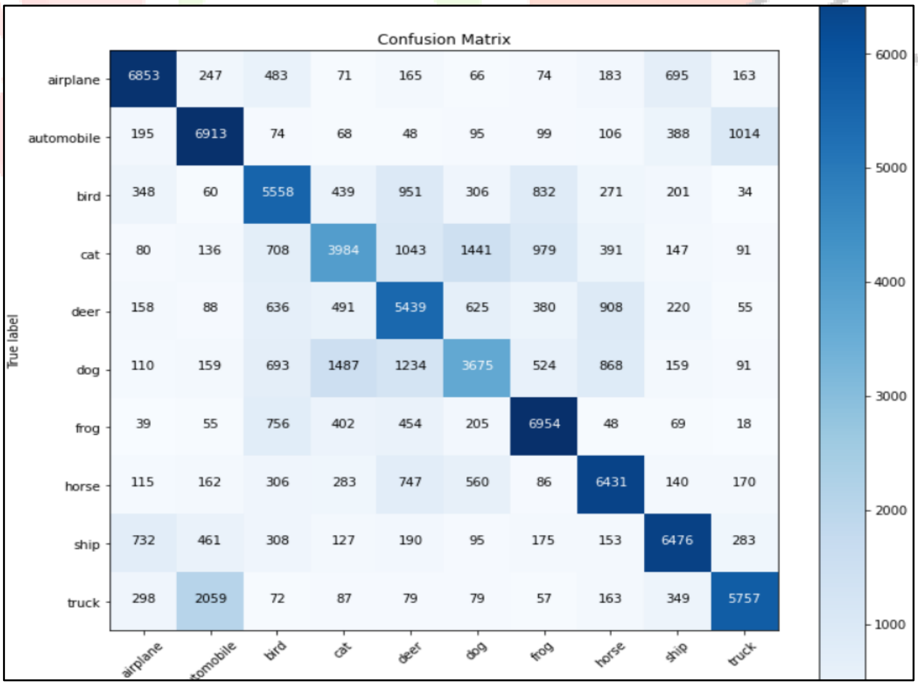

Figure 11 VGG16 Model Confusion Matrix

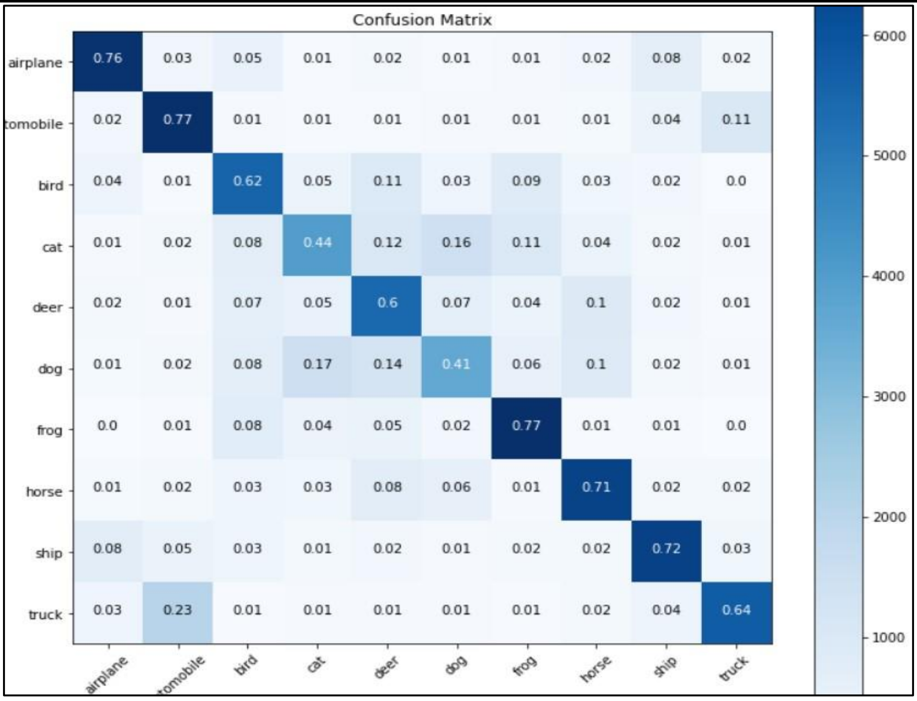

Figure 12 Accuracy of VGG16 Model

6.Conclusion

In summary, deep learning's incorporation into image retrieval systems represents a significant development in the processing of visual data. Sophisticated neural network topologies have demonstrated proficiency in extracting significant information from photos, improving the efficiency and accuracy of retrieval. Notwithstanding the achievements, issues with interpretability, scalability to big datasets, and adaptation to various picture formats still exist. Subsequent investigations ought to enhance current models, investigate innovative structures, and incorporate supplementary modalities to ensure system resilience. With the further development of deep learning and the integration of technologies like as attention mechanisms and transfer learning, it is possible to surmount present constraints and usher in a new age of smart, precise, and adaptable visual information retrieval systems.

7.References

[1]Jain, S., & Dhar, J. (2017, August). Image based search engine using deep learning. In 2017 Tenth International Conference on Contemporary Computing (IC3) (pp. 1-7). IEEE.

[2]Diana, M., Amagasaki, M., & Iida, M. (2020, November). Image Search System Based on Feature Vectors of Convolutional Neural Network. In 2020 IEEE REGION 10 CONFERENCE (TENCON) (pp. 934-939). IEEE.

[3]Fu, R., Li, B., Gao, Y., & Wang, P. (2016, October). Content-based image retrieval based on CNN and SVM. In 2016 2nd IEEE international conference on computer and communications (ICCC) (pp. 638-642). IEEE

[4]Agrawal S, Chowdhary A, Agarwala S, Mayya V, Kamath S S. Content-based medical image retrieval system for lung diseases using deep CNNs. Int J Inf Technol. 2022;14(7):3619-3627.

[5]Karthik K, Kamath SS (2021) A deep neural network model for content-based medical image retrieval with multi-view classifcation. Vis Comput 37(7):1837–1850.

[6]Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2022). Image retrieval based on deep learning. J. Syst. Manag. Sci, 12(2), 477-496.