JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 07 Issue 01
2024

Research Article
https://doi.org/10.56979/701/2024

Architectural Formation with Deep Learning and Algorithmic Bindings for Cross-Domain Information Retrieval

Khawaja Tehseen Ahmed[1], Muqadas Fatima[1*], Shahida Ummesafi[1], Aiza Shabir[1], Nida Shahid[1], Muhammad Yasir Khan[2], Ayesha Rubab[1], and Aleema Sadia[3]

[1]Department of Computer Science, Bahauddin Zakariya University Multan, Pakistan.
[2]Department of Computer Science, MNS University of Agriculture Multan, Pakistan.
[3]Department of Information Technology, Bahauddin Zakariya University Multan, Pakistan.
*Corresponding Author: Muqadas Fatima. Email: muqadasfatima988@gmail.com

Received: March 01, 2024 **Accepted:** May 21, 2024 **Published:** June 01, 2024

Abstract: Efficient strategies for index search are crucial elements involved in categorizing and retrieving simple as well as complex image collections and libraries. In this paper, new algorithm is presented aimed at refining the selection of images to be clustered and more accurate identification of ROIs in many clustered objects. The relations to other features are also expected to be provided, including the RGB image features and the other feature sets obtained with the use of Convolutional Neural Networks (CNNs) for achieving the scale invariance. Despite, GoogleNet and AlexNet and ResNet exist this algorithm has the deep feature and spatial data point of view for improving the image classification. Feature coefficient computation further enables the application of norms L1 and L2 on over the images of RGB. The 'Scale invariance' encompasses predicting the scaling of keypoints, computation of coefficients between two successive octaves along with expressions of virtual intra octave expressions. In the process of maxima selection, interpolation, non-maxima suppression, and cumulative thresholding the algorithm applies ROI detection. The presented multimodal approach significantly enhances the identification of objects particularly in a setting as depicted in this paper with high density of other similar objects. The color feature sets and CNN feature sets that are integrated in constructing the Bag-of-Words (BoW) model improve image indexation and image search. From the quantitative analysis, there is promising average precision (AP) and average recall (AR) when the presented algorithm is tested using data from Corel-10K, Tropical Fruits and Cifar-10 datasets.

Keywords: Convolutional Neural Network, Image retrieval,Bag of word, Cross-domain retrieval.

1. Introduction

The growing significance and application of digital media in various domains and industries due to its extensive use have resulted in increasing demands for new approaches to data representation and retrieval. Due to advancements in extracting relevant images out of extensive image databases, substantial enhancements have been made in existing image retrieval algorithms. Convolutional Neural Networks (CNN) designs have been integrated in order to enhance the efficiency of image search [1, 2]. In addition, deep learning-based feature extraction methods have also improved the multimedia content processing. The CNN features [3] have been proved to be very effective in many retrieval and computer vision tasks and make CNN a fundamental model in deep learning. Deep learning, which focuses on the complex construction of sophisticated artificial neural networks for feature extraction and classification tasks, is

exceptionally suitable in image retrieval because of its superior ability to understand the texture or complex spatial patterns and unique recognition features. Deep learning also helps in feature detection and understanding within an image. In combination with Content Based Image Retrieval (CBIR), it creates a robust approach for mediating meaning of the visual content meaningfully. The type of visual features used in a CBIR system determines the efficiency and preciseness of the process of searching and retrieving the content. Visual features can be broadly classified as low-level features that are inclusive of shape [4], color [5, 6] spatial relationship and texture [7, 8] and high-level features which can also be described as semantic features. CBIR concerns itself with the relationship between the spatial arrangement of the low-level features and the semantic concepts [9].It is therefore very important for the understanding of images and the interpretation of semantic context of the images for improving precision and efficiency of image retrieval systems. CBIR is generally based on the process of deriving features from a database of saved images. When a user submits a query image for search, the CBIR system details like color, shape and texture are extracted and compared to the feature database to retrieve relevant images. In this work, a novel approach to the smart image search is presented with an emphasis on analyzing areas of potential object density in images. RGB image features are integrated with CNN extracted features for scale invariance, and it uses the sequential fusion of deep features of GoogleNet, AlexNet, and ResNet for better detection. These often involve computing RGB feature coefficients, careful and precise key point scaling and using multiple strategies for the identification of regions of interest. The algorithm significantly enhances object detection especially in complex structures using color and CNN features to perform image indexing and retrieval. Examining across mixed datasets proves its efficiency regarding the Average Precision and Recall metrics.

Following are some of the contributions:

A novel algorithm for enhancing image selection and ROI identification within a small category of objects is presented.

- The scales of the images are enabled by the RGB image features and the features extracted from the CNN.
- Provides better image classification achieved by using the L1 and L2 norm and key points scaling results with help of the Extended GoogleNet, AlexNet, and ResNet that are applied for deep feature integration and spatial data perspective.
- Improved object detection achieved through maxima selection, interpolation, non-maximum suppression and cumulative thresholds techniques in multistage.
- Performs intersections between the color feature sets and BoW model with the CNN feature sets to enhance the image indexing and the search ability.

The higher AP and AR recall rates provided by Corel-10K, Tropical Fruits, and Cifar-10 enlighten effectiveness of the presented algorithm.

The subsequent sections of the article are structured as follows: Section 2 is about the literature review where the focus is set on CNNs within the framework of deep learning. The research method that is used is described in section 3. Section 4 contains the analysis of research findings presented with the help of tables and graphs. Lastly, Section 5 provides a conclusion based on the provided methodology.

2. Materials and Methods

The approach suggested in [10] aims at expanding the concept of using spatial color data and shape-based features, as well as object recognition for the improvement of information fusion. This entails mapping L2 spatial color configurations for RGB channels and correlating shape features from the grey scale images depending on intensity. In the first step, feature vectors are reduced, and in the second step, only the coefficients with the highest variance are chosen; as for the image indexing and retrieval process, a Bag-of-Words approach is used. In images, features are detected [11] from the interest points obtained through the non-maximum suppression of derivatives. These points are then described using scale space approaches and are finally fused with color information. Small values of data size mean that high variance coefficients are chosen. The article [12] sheds light on the implementation process of deep learning algorithms like GoogleNet, VGG-19, and ResNet-50 in the image processing domain. In this paper, they talk about their approach involving the use of few connections and multiple layers for discriminative mapping. The study incorporates these architectures with textured Eigenvalues and object features

standardized by convolutional Laplacian, in addition to the mapped color channels to ultimately enhance image retrieval proficiency with regard to different extent and benchmarks of semantic categories. The model emphasizes the efficiency in time and computational treatment, enhancing deep learning fusion and new descriptor design. Using a content-based approach presented in [13] with the deep convolutional neural networks (CNNs), one can achieve high accuracies in image retrieval. This involves the integration of GoogleNet and VGG-19 in a manner that results in a formation of signatures that contains the texture, color and shape of objects. First, by Markov Random Field (MRF) classifier, the maximum response of the texture patterns is obtained. Following this, key points are detected using the Fast Retina Key point (FREAK) descriptor that works by detecting corners instead of edges. GoogleNet and VGG-19 models are employed to extract deep features, and correlogram method to calculate color attributes. The derived methodology [14] offers a method for detecting regions of interest inside images as well as giving innovations to them in the global domain. Several interest points are accumulated over multiple representation levels to generate signatures of the image. First, the shape features are obtained by grouping connected pixels through employing a binary intensity thresholding. Finally, features inside regions of maximum stability are localized using the histogram of oriented gradients (HOG). These features are combined with rotation-invariant texture characteristics obtained from uniform local binary patterns (LBP) as well as rearrangement. This algorithm links the region and texture into fewer components which assist in computation of principal components. Image features as objects and color are extracted and described for analysis [15].These methods include Gaussian of variance and image convolution; matrix hashing for feature approximation. It also divides high-range frequency feature information for spatial indexing, which is effective when implementing the strategy for high resolution images with various semantic segments. Grey-level details are united with features based on RGB channels in the method of principal component analysis, statistical information is minimized. Lastly, for image indexing and retrieval, the Bags-of-Word setup is employed. In [16], a feature descriptor is proposed that performs interest point detection on a sliding window that is fixed to an optimal scale and examines textures on the corresponding patterns. It builds up on the Moravec technique by using the covariance matrix derived from local first order derivatives to detect corners, edges or noise elements. These potential points of interest are then passed through sliding window approach to extract strong features. The method presented in [17] has pointed out the significance of symmetry across the different image processing techniques being performed over the CNNs, namely sampling, scoring, scaling, filtering, and suppression. The method inclusions involve; symmetric sampling around key points, rotated sampling patterns and smoothing of the image by using standard deviation. PCA is then applied to the ResNet features after the features have been extracted and scaled followed by a concatenation of the central features.

A new method is presented in [21] that belongs to the type of CBIR and utilizes support vector machine, K-nearest neighbors and Convolutional Neural Networks. Deep learning techniques for low and high level image features are illustrated by using Corel 1K, 5K, and 10K databases in the research. In the present emphasis, the findings show that deep learning models enhance the performance and precision of image related search activities.

Another work on CBIR proposed in [22] is a two-part hybrid deep machine learning approach. Applying transfer learning with ResNet50 and VGG16, and then utilize KNN to match similar features, the work attains up to 97.4% image search accuracy by searching through large databases of data. The approach has significant enhancements in some areas including the use of digital libraries and cases of crime control and finger print identification.

A new method namely RetCCL is proposed in [23] where clustering-guided contrastive learning is applied for WSI retrieval. This method is beneficial because it generates different patches that embrace the whole image hence enhances a better accuracy and speed in recovering the damaged areas. The method proves especially useful for medical image databases, where the correct results and the ability to handle large numbers of images are fundamental to the application.

Specifically, a transformer-based hashing framework named VTDH is introduced in [24] that integrates neural networks and hashing approaches for enhancing the image search quality. The study indicates that the suggested framework is more accurate and much faster than the existing approaches on CIFAR-10 and ImageNet databases.

The survey in [25] and discusses various types of deep learning based retrieval algorithms. A key feature of this work is how it reveals how the use of these algorithms increases the effectiveness and precision of recommendations through the effective identification of most suitable images that will meet users' search queries. There is a new method called Centralized Spherical Quantization (SCQ) made in [26] which is aimed to improve the time of image search. Thus, by improving the quantification process, SCQ achieves up to 85% higher map than current algorithms in CIFAR-10 and NUS-WIDE datasets. A CBIR system that is elaborated in [27] [28] that combines feature fusion techniques which includes edge directions, texture features, color histogram, and deep learning models. The study shows the efficiency of the described system for historical investigation and architecture design, as well as for targeted image search that provides high accuracy.

2.1. Methodology

2.1.1. Detection of Region of Interest

Scale invariance is an important feature that helps the algorithms to detect the similar key points in images even if the size of the stored objects is different. Object detection of key points is therefore mandatory, whether the object is large or small or even close up or far away. In this stage, basic characteristics of the input are obtained and possible areas are recommended. These regions are then classified or segmented based on their content it would be an effort to recognize areas within the data that are important so as to be used for further processing. To accomplish this, methods like feature extraction, region proposal, and classification are used for precise location of the target irrespective of the size or appearance of regions of interest [30].

2.1.2. Scale Spaces

In image processing, scale invariance is done by converting the image into a scale space representation. This process entails constructing multiple octaves of the image; each octave being the scaled-down version of the original image. Octaves are created by continuously subdividing the image and applying a smoothing filter to them before down-sampling. Down sampling helps in the reduction of dimensions of the images and enhances the efficiency of processing them. Every octave contains the same image but seen at successively higher resolutions to show details at different scales. The process of down sampling is then performed successively for each octave to form image pyramids with different levels that provide different scales. It guarantees that the key points can be detected for the object of interest independently from its size within the image and across the different scales. For this reason, it is possible to assist algorithm gain scale invariance through the representation scale-space, as it is possible to detect key points in scale levels in the image [31] [32]. It becomes possible to make a relevant matching and recognizing in several angles of the object detection, the image registration and the search and find of images and the problem occurs when the objects are of different sizes and distances from the camera.

2.1.3. Masking

Mask 9-16 is used in signal processing for the purpose of filtering high gross noise and enhancement of image quality. It works using 3×3 mask placed on the image and each element in the mask is an average of the pixels in the surrounding region weighted. They are determined by the Gaussian function in which the central sample or pixel is given the highest weight and the succeeding samples or pixels are granted progressively lower weights. This process of smoothing the image reduces the random variation and also improves the sharpness and outlines of the image by preserving the important features. Circular neighbor 8 is one of the common techniques in feature extraction in the field of image processing to obtain circular objects or structures of the image [32]. It is one of the steps in feature extraction process that helps in the detection of objects and categorizing them within the imaginable scene. Circular neighbor 8 divides the image into several scales and subtract an unwanted noise in the image using any of the smoother masks 9-16.It is the third dimension of feature extraction, and its outcomes are integrated with other techniques to improve chances for object identification in the image.

2.1.4. Maximum Threshold

Non-maximum suppression is used to further refine the selection of the corner points when the corner detection is being done. This process measures the significance of each of the corner points defined in an area or the nearby area that is a radius or window length from the point in consideration. If the corner point being evaluated is the maximum within its local neighborhood, it means that it will be included as a candidate corner. However, if it is not the maximum, it is suppressed or discarded. This helps in the

removal of some or most of the corner points which are less relevant or unnecessary hence retaining the dominant corner points.

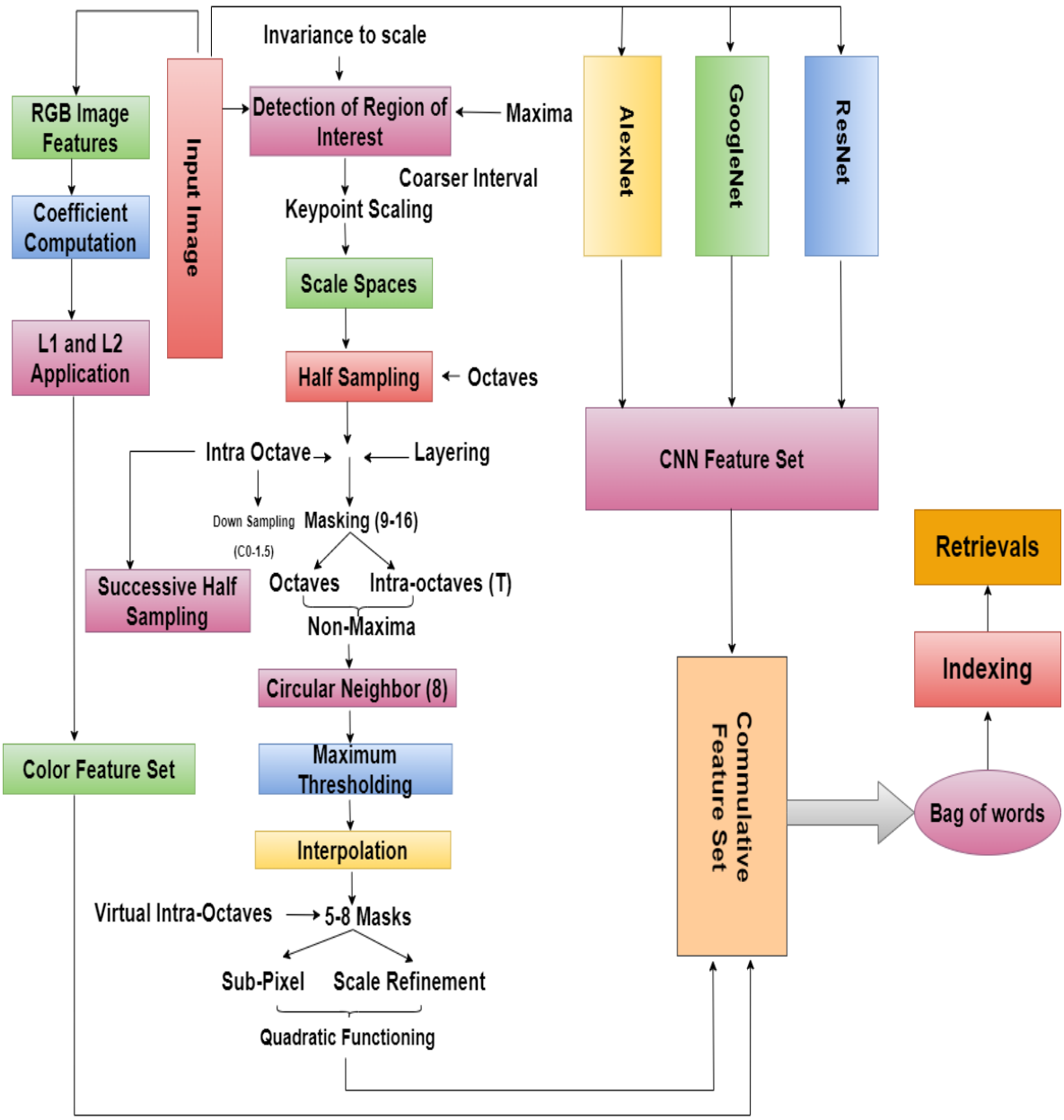


Figure 1. Proposed Methodology

2.1.5. Non Maxima Suppression

The application of non-maximum suppression by choosing only the important corner points leads to better accuracy and reliability of corner detection algorithms and their insensitivity to noise and variation in images. These identified corner points can be further used for various applications including feature matching, object recognition or image registration. Interpolation is used to determine the unknown response value when the pixel positions are not distinct or not on the response surface. They are used in the non-maxima suppression algorithm at patch boundaries to maintain a proper flow of responses while sustaining the maximum condition.

2.1.6. Interpolation

This interpolation procedure helps to select the more significant key points that would capture unique and different patterns at different scales. When key points have been detected, their precise location is then determined more accurately with the help of sub-pixel precision. In this process, the detail of the lightness gradient present at the key point location and the pixel value interpolation to decide the appropriate sub-pixel position is performed. The gradient information is then used to find the direction and magnitude of

the local image structure around the key point, and these features are used along with the intensity values to better refine the location of the key point. When the location of the key point is refined to sub-pixel accuracy, the step of feature matching and recognition improves significantly in terms of resolution and quality since it can effectively distinguish key points that are very close or those that look very similar.

2.1.7. Color Feature Set

First, the image is described as a set of key points that are decisive points to recognize the image. Coefficient computation is an approach applied in image scaling to maintain the colors of an image as the size is adjusted. Here the image is implemented as a two-dimensional matrix of pixels where each pixel has a certain color value. The color values are analyzed by a color feature set containing a set of mathematical functions that represent the color information of the image. These functions are used in the evaluation of coefficients that correspond to each pixel in the image. These coefficients are then used to scale the dimensions of the image while at the same time preserving the color data measurements. This is done in scaling where each pixel color value is multiplied with its corresponding coefficient depending on the scale factor to change the color of the pixels. This type of scaled image is also pretty capable of preserving the color details of the original image and has easily recognizable colors even after scaling. This is particularly so especially in activities such as object recognition similar to image search where color data is essential in the identification and matching processes.

2.1.8. CNN Feature Set

The proposed method includes deep models such as the deep convolutional neural network which includes models like AlexNet, GoogleNet, and ResNet. CNN is one of the most powerful and efficient approaches in the field of deep learning, which is associated with image analysis. This is possible because through the filter, features may be analyzed and segmented from an image; something that makes it ideal when developing image processing systems. Some of the main parts that make up CNN include the following: In the architecture of the proposed model the first layer is the convolutional layer, the second layer is the pool layer, the third layer is the fully connected layer and the fourth layer is the output layer. The convolution layer's major role is to transform an input image into features where in this layer there are filters that are passed through the image. These filters apply convolution with the input data to generate feature maps. Regardless of max-pooling or average-pooling, pooling layers help to achieve dimensionality reduction of feature maps as well as prevent overfitting. Dense layers or fully connected layers are the final layers in a neural network model that make the final prediction based on extracted features. They are another set of neurons that receive input from the previous layer and apply a linear operation to provide the final result. The output layer as often a Softmax layer that gives each class label a probability so that the network can make its decision. AlexNet has eight layers out of which, five layers are convolutional layers meant to define the features while the other three are fully connected layers that help to carry out the classification. On the other hand, GoogleNet comes with 22 layers of the convolution and fully connected layers and has adopted the new "Inception module" which can work on multiple feature maps at the same time and still minimize on the network parameters but at the same time maximizing on the accurate models. The idea introduced by ResNet is called residual connections, which directly allow gradients to pass between layers and solve the vanishing gradient problem to enhance deep learning. Architecturally, it is made up of several residual blocks placed in a linked manner, featuring convolutional layers, batch normalization and ReLU activation, with skip connections to enable efficient training and reduce the computational complexity of learning residual as well as non-residual information.

2.1.9. Commutative Feature Set

The system consists of a feature set known as CNN feature set; this component is responsible for object detection and classification within an image. The CNN feature set is related to the Intra-octaves layering and the Maxima feature set which is very useful in processing the image and determining maximum features from the image. The components of the system are integrated so that it can accept images of different formats and dimensions with optimum speed and accuracy.

2.1.10. BoW Architecture

The bag-of-words model, which is a fundamental component of NLP, converts textual data into vectors of numerical values based on the count of words. It builds the vocabulary list for all the documents in the corpus resulting into high dimensions and low-density vectors, where the dimensions represent words and values representing word frequencies. N-grams are disregarding word order and found useful

in several NLP applications such as sentiment analysis, document classification, and information retrieval where it often serves as features for machine learning.

2.2. Experimentation

2.2.1. Datasets

The effectiveness of an image retrieval system is measured on the basis of several datasets containing different images attributes like color, texture, and shape. This assessment is designed in such a way that guarantees that it is both efficient and accurate. The experiment carried out in this study makes use of three datasets comprising of different categories; Cifar-10, Tropical Fruits and Corel-10K. The datasets consist of images collected from different sources, thus the evaluation of many features of images is possible. To be efficient and accurate at the same time, the experiments are carried out across a range of datasets. These numerical outcomes for each set of data show the varying degrees of accuracy and efficiency for each set of data based on its characteristics.

2.2.1.1 Cifar-10

The Cifar-10 dataset contains sixty thousand color images each of size 32×32 categorized into ten classes with six thousand images in each class. It is extensively used as a training and testing dataset in deep learning, particularly for image classification applications. Images in the database are in RGB format and each of them belongs to one of the ten classes: automobile, birds, ships, deer, dogs, frogs, cats, horses and trucks. The images are distributed evenly across each class, and each class contains approximately 8,000 while the background class contains nearly 6,000 images. This particular dataset is widely used as a benchmark to measure the performance of image classification models.

Figure 2. Cifar-10 dataset showing different sample images belong to each category [18]

2.2.1.2. Corel-10K

The Corel-10K dataset is a vast compilation comprising 100 distinct categories, each housing 100 individual images, totaling 10,000 images (100 x 100). Every image has dimensions of 128 × 84 pixels. The dataset is primarily intended for research and experimentation in image retrieval and feature extraction techniques.

Figure 3. Corel-10K dataset showing different sample images belong to each category [19]

2.2.1.3. Tropical-fruits

The tropical fruits dataset is a comprehensive compilation featuring fifteen categories of fruit, encompassing both fresh and rotten variations. The dataset comprises 3,200 original images and 12,335 augmented images, initially trained on a set of 2,560 images.This database is divided into 15 categories i.e. asterix_potato, granny_smith_apple, nectarine, plum, watermelon, kiwi, honneydew_melon, cashew, onion, Spanish_pear, ta iti_lime and diamond_peach.

Figure 4. Sample images are selected from respective tropical-fruits dataset [20]

2.2.2. Precision, Recall and F-measure Evaluation

Precision and recall play pivotal roles in assessing the effectiveness of information retrieval systems. Precision evaluates the accuracy of retrieved results by measuring the ratio of relevant items to the total retrieved set. A high precision value indicates a strong alignment between retrieved items and user queries, while a low precision value suggests the presence of irrelevant data. Precision can be computed as:

$$Precision = \frac{N_{q(a)}}{N_{p(b)}} \quad (1)$$

Recall assesses how well a system captures all relevant items, indicating the percentage of correctly retrieved relevant items out of the total relevant items. High recall reflects efficient retrieval of a large portion of relevant items, while low recall indicates the system's failure to retrieve a significant number of relevant items, leading to incomplete outcomes. Recall can be computed:

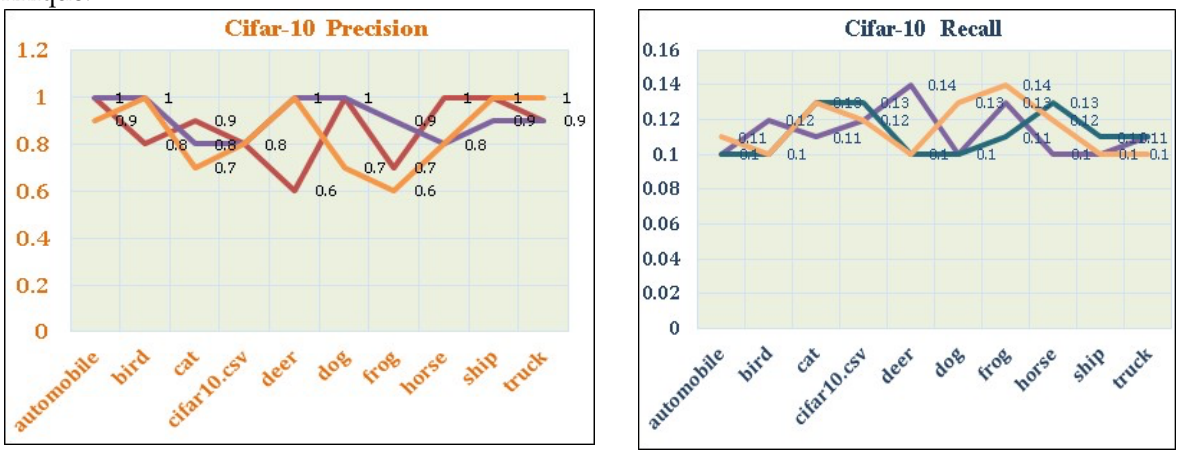
$$Recall = \frac{N_{q(b)}}{N_{\circ}} \quad (2)$$

The F-Measure is a reliable method that combines precision and recall into a single measurement, providing a comprehensive assessment of performance by skillfully incorporating both factors. The F-measure can be computed as:

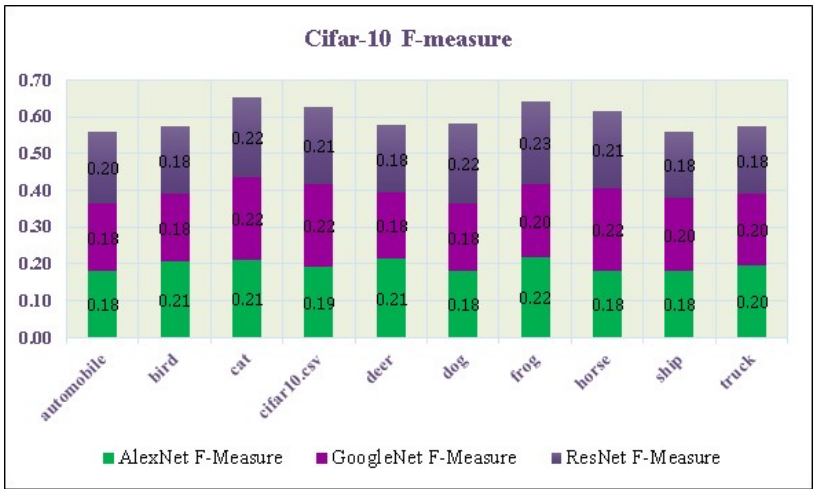
$$F = \frac{2*m*n}{m+n} \quad (3)$$

3. Results

We conducted experiments on various benchmark datasets to evaluate the accuracy of the proposed technique.



(a) Precision **(b) Recall**



(c) F-Score

Figure 5. (a) Show the Average Precision, (b) Show the Recall, (c) Show the f-score of Cifar-10 dataset is with CNN Architecture

Figure 5a presents the average precision rates for the CIFAR-10 dataset using CNN architecture. The suggested approach demonstrated an average precision ratio exceeding 90% for image categories like automobiles, dogs, ships and horses, while attaining average precision results exceeding 85% forcatand truck. Furthermore, the offered method demonstrated average precision rates exceeding 70% for various additional categories. The frog category shows 0.14 recall rate with ResNet. The bird category shows highest recall rate 0.12 with AlexNet. The horse category shows 0.13 recall rate with GoogleNet. ResNet architecture scores highest F-measure value of 0.23. Extreme value for AlexNet and GoogleNet is 0.22 and most of the categories the F-measure value is 0.18.

Table 1. Precision, recall, and F-score for the Cifar-10 dataset using AlexNet, GoogleNet, and ResNet

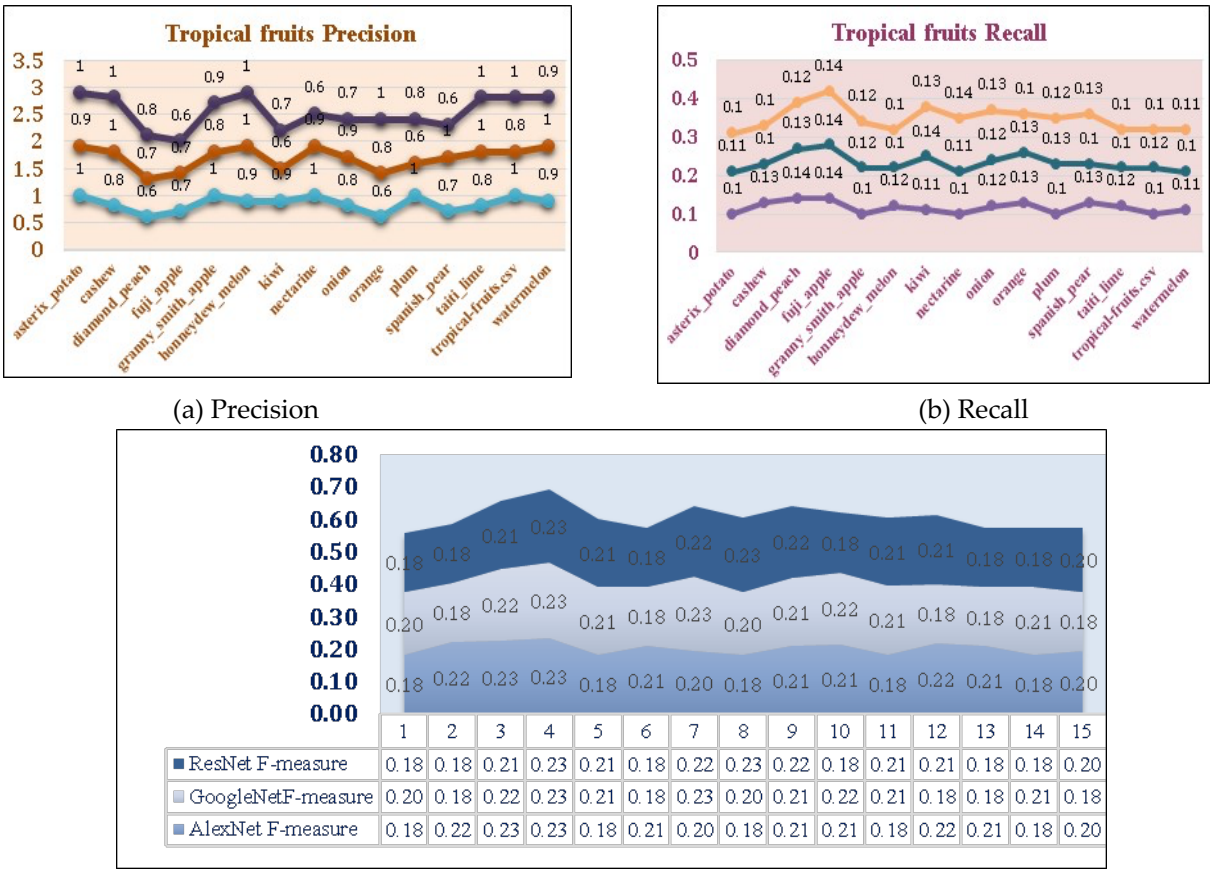
Cifar-10 Dataset									
	AlexNet			GoogleNet			ResNet		
Category	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-score**
automobile	1	0.1	0.18	1	0.1	0.18	0.9	**0.11**	0.20
Bird	0.8	**0.12**	0.21	1	0.1	0.18	1	0.1	0.18
cat	0.9	0.11	0.21	0.8	0.13	0.22	0.7	0.13	0.22
cifar10.csv	0.8	0.12	0.19	0.8	**0.13**	**0.22**	0.8	0.12	0.21
deer	0.6	**0.14**	0.21	1	0.1	0.18	1	0.1	0.18
Dog	1	0.1	0.18	1	0.1	0.18	0.7	**0.13**	0.22
Frog	0.7	0.13	**0.22**	0.9	0.11	0.20	0.6	**0.14**	**0.23**
horse	**1**	0.1	0.18	0.8	**0.13**	0.22	0.8	0.12	0.21
Ship	1	0.1	0.18	0.9	**0.11**	0.20	1	0.1	0.18
Truck	0.9	0.11	0.20	0.9	0.11	0.20	**1**	0.1	0.18

3.1. Experiments on Tropical-fruits dataset

The proposed approach processes all categories of tropical fruits dataset. Furthermore, the proposed approach excels in certain categories such as asterix_potato, granny_smith_apple, nectarine, plum, watermelon, kiwi, honneydew_melon, cashew, onion, Spanish_pear, ta iti_lime and diamond_peach.

The suggested approach demonstrated an average precision ratio exceeding 90% for image categories like asterix_potato,, cashew, granny_smith_apple,watermelon, honneydew_melonand onion while achieving above 70% average precision results for diamond-peach, nectarine and Spanish-pear. Furthermore, the fuji-apple and nectarine categoryshow0.14 recall rate with ResNet. The diamond-peach category shows highest recall rate 0.14 with AlexNet. Thefuji-apple and kiwicategory show0.14 recall rate

with GoogleNet. The extreme F-measure value for AlexNet, GoogleNet, and ResNet architectures is 0.23, while for most categories, the F-measure value is 0.18.



(a) Precision (b) Recall

(c) F-Score

Figure 6. (a) Illustrates the Average Precision, while (b) presents the Recall, and (c) displays the F-score for the CIFAR-10 dataset utilizing AlexNet, GoogleNet, and ResNet architectures

The suggested approach demonstrated an average precision ratio exceeding 90% for image categories like asterix_potato,, cashew, granny_smith_apple,watermelon, honneydew_melonand onion while achieving above 70% average precision results for diamond-peach, nectarine and Spanish-pear. Furthermore, the fuji-apple and nectarine categoryshow0.14 recall rate with ResNet. The diamond-peach category shows highest recall rate 0.14 with AlexNet. Thefuji-apple and kiwicategory show0.14 recall rate with GoogleNet.The extreme F-measure value for AlexNet, GoogleNet, and ResNet architectures is 0.23, while for most categories, the F-measure value is 0.18.

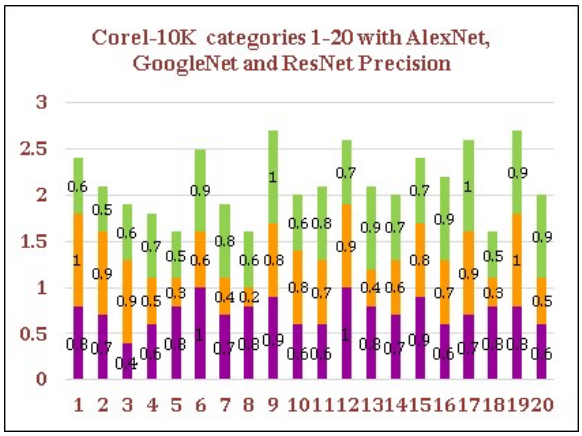
Table 2. The Precision, Recall, and F-score metrics of AlexNet, GoogleNet, and ResNet with the Tropical-fruits dataset

Tropical-fruits Dataset									
	AlexNet			**GoogleNet**			**ResNet**		
Category	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
asterix_potato	1	0.1	0.18	0.9	0.11	0.20	1	0.1	0.18
cashew	0.8	0.13	0.22	1	0.1	0.18	1	0.1	0.18
diamond_peach	0.6	**0.14**	**0.23**	0.7	0.13	0.22	0.8	0.12	0.21
fuji_apple	0.7	**0.14**	**0.23**	0.7	**0.14**	**0.23**	0.6	**0.14**	**0.23**
granny_smith_apple	1	0.1	0.18	0.8	0.12	0.21	0.9	0.12	0.21

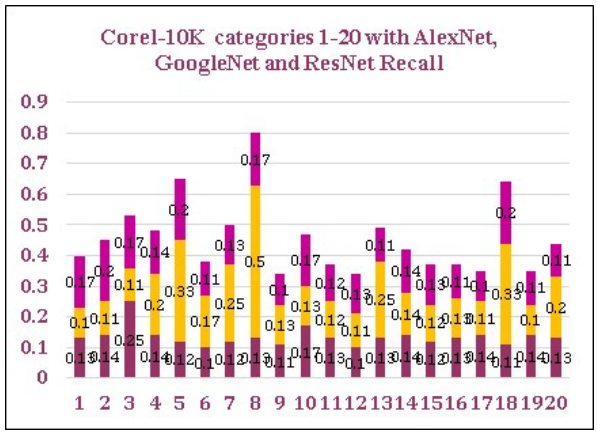
honneydew_mel on	0.9	0.12	0.21	1	0.1	0.18	1	0.1	0.18
kiwi	0.9	0.11	0.20	0.6	**0.14**	**0.23**	0.7	0.13	0.22
nectarine	1	0.1	0.18	0.9	0.11	0.20	0.6	**0.14**	**0.23**
onion	0.8	0.12	0.21	0.9	0.12	0.21	0.7	0.13	0.22
orange	0.6	0.13	0.21	0.8	0.13	0.22	1	0.1	0.18
plum	1	0.1	0.18	0.6	0.13	0.21	0.8	0.12	0.21
spanish_pear	0.7	0.13	0.22	1	0.1	0.18	0.6	0.13	0.21
taiti_lime	0.8	0.12	0.21	1	0.1	0.18	1	0.1	0.18
tropical-fruits.csv	1	0.1	0.18	0.8	0.12	0.21	1	0.1	0.18
watermelon	0.9	0.11	0.20	1	0.1	0.18	0.9	0.11	0.20

3.2. Experiments on large benchmark Corel-10K dataset

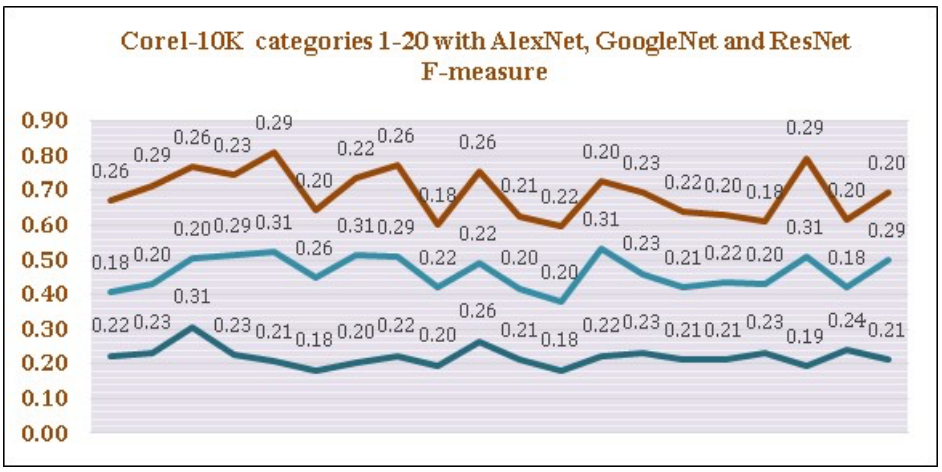
The Corel-10K dataset contains 10,000 images categorized into 100 diverse categories, with each category comprising 100 individual images.



(a) Precision



(b) Recall



(c) F-Score

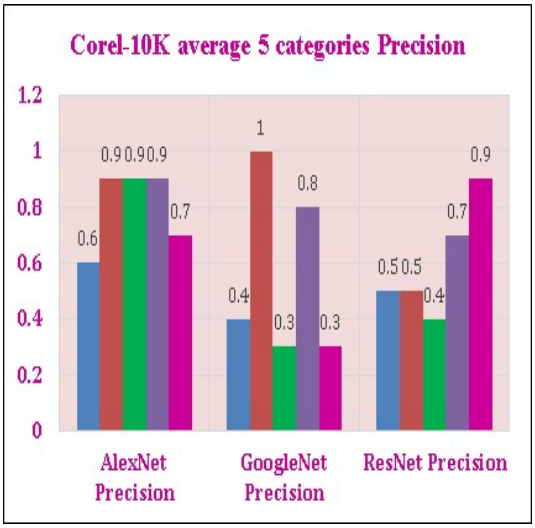
Figure 7. (a) Shows the Average Precision, (b) shows the Recall, and (c) shows the F-score of top 20 categories ofCorel-10K dataset with AlexNet, GoogleNet, and ResNet architectures

The graphical representation depicted in Figure 7illustrates the top 20 categories within the Corel-10K dataset, complemented by their corresponding numerical values as detailed in Table 3. Categories 14 and 2 exhibit the highest average precision of 1.0 when assessed with AlexNet, while Categories 10 and 26 achieve the same precision score with GoogleNet. Similarly, Categories 17 and 24 demonstrate a perfect average precision of 1.0 when evaluated with ResNet. Notably, Category 8 achieves the highest recall score of 0.5 within the framework of the GoogleNet architecture. Furthermore, Categories 16 and 27, analyzed

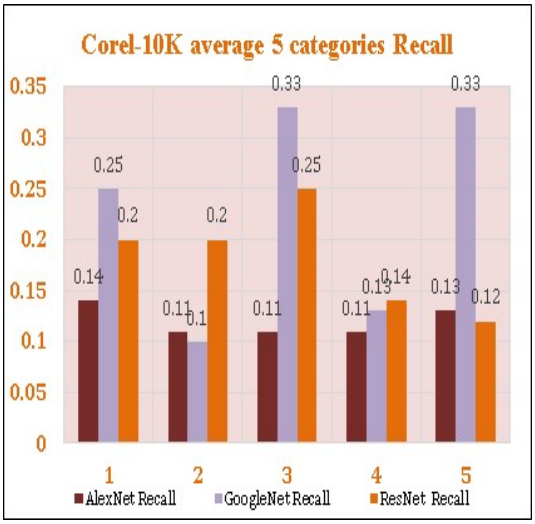
with GoogleNet, yield the highest F-score of 0.29, whereas Categories 100 and 25, examined using ResNet, also achieve this same F-score.

Table 3. Precision, Recall, and F-measure scores for AlexNet, GoogleNet, and ResNet on the top 20 categories of the Corel-10K dataset

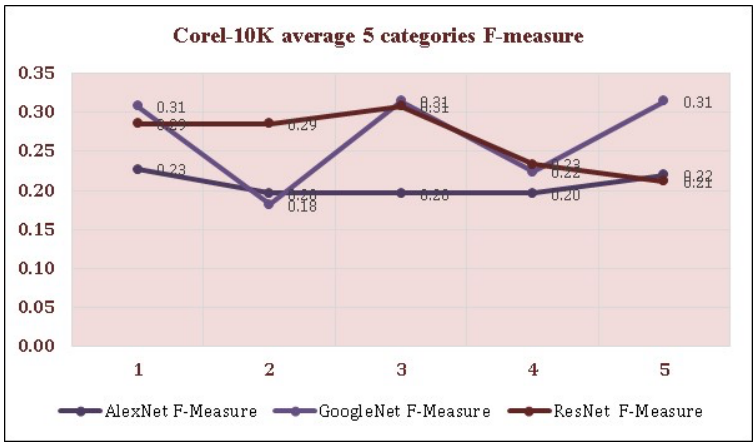
Corel 10K dataset top 20 categories									
	AlexNet			GoogleNet			ResNet		
Category	Precision	Recall	F-score	Precision	Recall	F-score	Precision	Recall	F-score
10	0.8	0.13	0.22	1	0.1	0.18	0.6	0.17	0.26
100	0.7	0.14	0.23	0.9	0.11	0.20	0.5	0.2	0.29
11	0.4	0.25	0.31	0.9	0.11	0.20	0.6	0.17	0.26
12	0.6	0.14	0.23	0.5	0.2	0.29	0.7	0.14	0.23
13	0.8	0.12	0.21	0.3	0.33	0.31	0.5	0.2	0.29
14	1	0.1	0.18	0.6	0.17	0.26	0.9	0.11	0.20
15	0.7	0.12	0.20	0.4	0.25	0.31	0.8	0.13	0.22
16	0.8	0.13	0.22	0.2	0.50	0.29	0.6	0.17	0.26
17	0.9	0.11	0.20	0.8	0.13	0.22	1	0.1	0.18
18	0.6	0.17	0.26	0.8	0.13	0.22	0.6	0.17	0.26
19	0.6	0.13	0.21	0.7	0.12	0.20	0.8	0.12	0.21
2	1	0.1	0.18	0.9	0.11	0.20	0.7	0.13	0.22
20	0.8	0.13	0.22	0.4	0.25	0.31	0.9	0.11	0.20
21	0.7	0.14	0.23	0.6	0.14	0.23	0.7	0.14	0.23
22	0.9	0.12	0.21	0.8	0.12	0.21	0.7	0.13	0.22
23	0.6	0.13	0.21	0.7	0.13	0.22	0.9	0.11	0.20
24	0.7	0.14	0.23	0.9	0.11	0.20	1	0.1	0.18
25	0.8	0.11	0.19	0.3	0.33	0.31	0.5	0.2	0.29
26	0.8	0.14	0.24	1	0.1	0.18	0.9	0.11	0.20
27	0.6	0.13	0.21	0.5	0.2	0.29	0.9	0.11	0.20



(a) Percision



(b) Recall



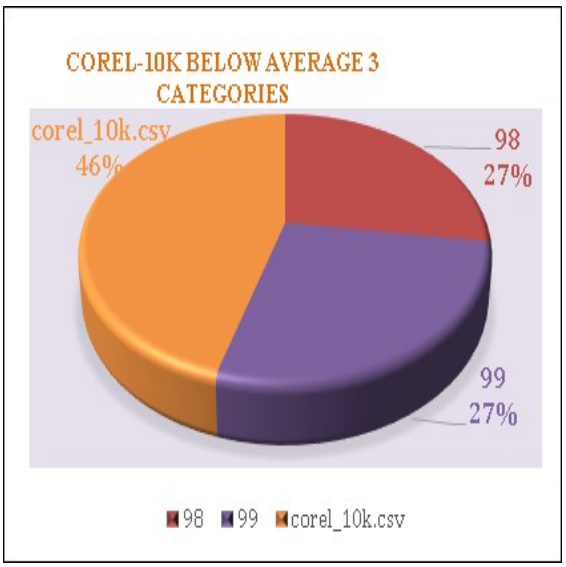
(c) F-Score

Figure 8. (a) shows the Average Precision, (b) shows the Recall, and (c) shows the F-score of average 5 categories of Corel-10K dataset with AlexNet, GoogleNet, and ResNet architectures

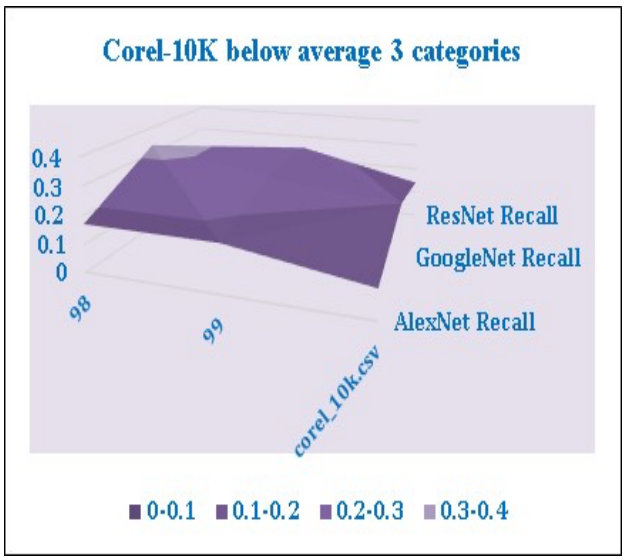
In Figure 8a, the Average Precision is depicted, while Figure 8b displays the Recall rate, while Figure 8c presents the average F-score across five categories from the Corel-10K dataset, analyzed across the AlexNet, GoogleNet, and ResNet architectures. Notably, Category 53 demonstrates 100% precision when evaluated with GoogleNet. Furthermore, the highest recall rate of 0.33 is observed for categories 54 and 56, utilizing the GoogleNet architecture. Categories 52 and 54 exhibit the highest F-score of 0.29 when assessed with the ResNet architecture. These numerical values are presented in Table 4.

Table 4. Average Precision, Recall, and F-measure scores across five categories for AlexNet, GoogleNet, and ResNet on the Corel-10K dataset

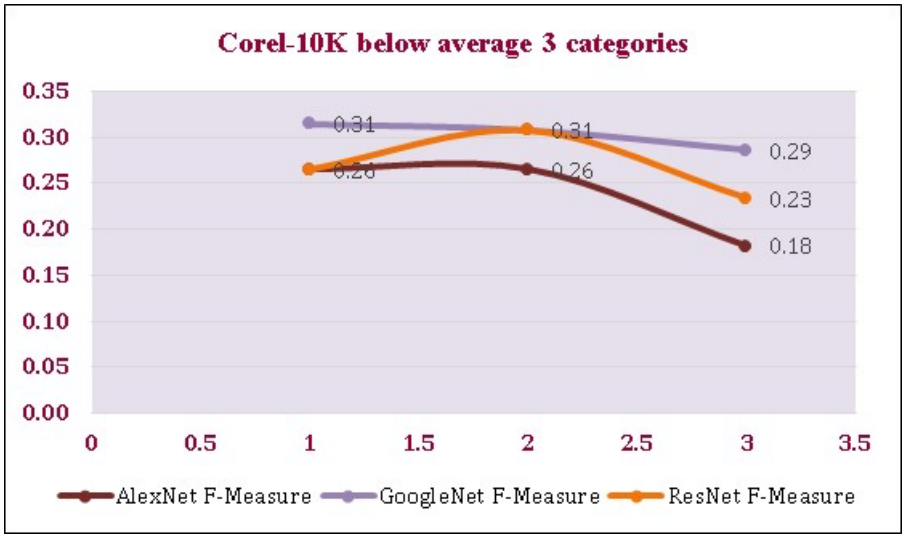
Corel 10K dataset average 5 categories									
	AlexNet			GoogleNet			ResNet		
Category	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
52	0.6	0.14	0.23	0.4	0.25	0.31	0.5	0.2	**0.29**
53	0.9	0.11	0.20	**1**	0.1	0.18	0.5	0.2	**0.29**
54	0.9	0.11	0.20	0.3	**0.33**	0.31	0.4	0.25	0.31
55	0.9	0.11	0.20	0.8	0.13	0.22	0.7	0.14	0.23
56	0.7	0.13	0.22	0.3	**0.33**	0.31	0.9	0.12	0.21



(a) Precision.



(Recall)



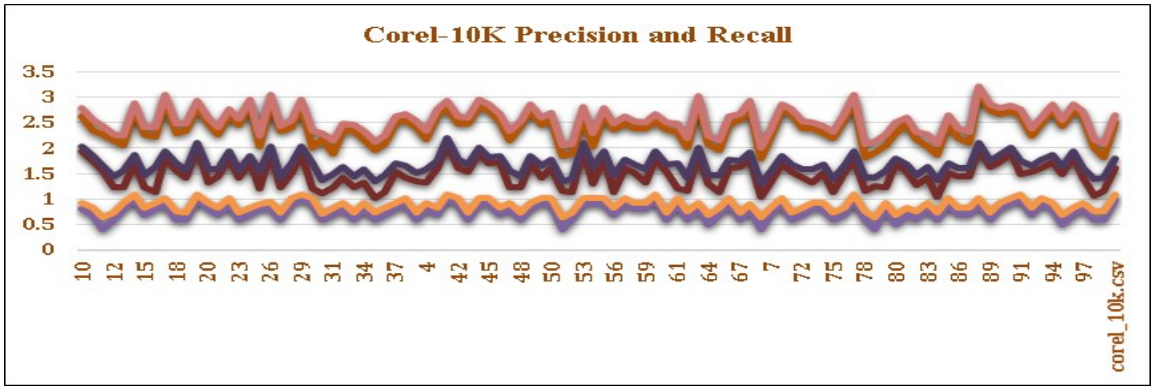
(c) F-Score

Figure 9. (a) shows the Average Precision, (b) shows the Recall and (c) shows the F-score of below average 3 categories of Corel-10K dataset with AlexNet, GoogleNet, and ResNet architectures

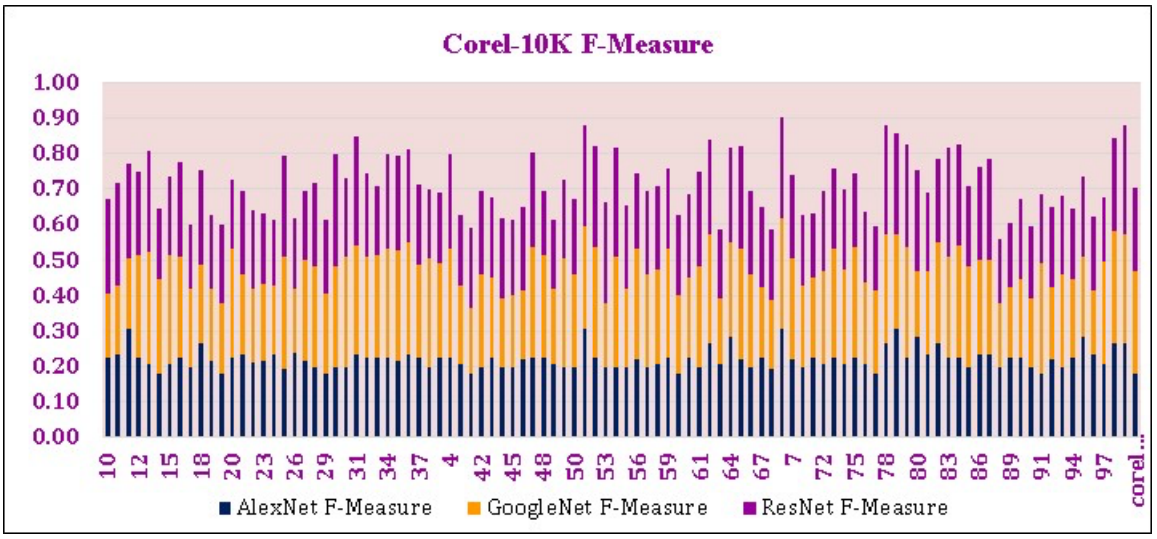
In Figure 9a, the Average Precision is depicted, while 9b illustrates the Recall rate, and 9c showcases the F-measure of the below average of 3 categories from the Corel-10K dataset, analyzed across the AlexNet, GoogleNet, and ResNet architectures. Noteworthy is the demonstration of 100% precision by Category corel-10K.csv when assessed with AlexNet. Additionally, the GoogleNet architecture showcases the highest recall rate of 0.33 for category 98. Furthermore, both Category 98 and 99 exhibit the highest F-score of 0.31, with GoogleNet and ResNet architectures, respectively. These comprehensive numerical insights are meticulously presented within the confines of Table 5.

Table 5. Precision, Recall and F-measure with AlexNet, GoogleNet and ResNet of Corel-10K dataset below average 3 categories

Corel 10K dataset below average 3 categories									
	AlexNet			**GoogleNet**			**ResNet**		
Category	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
98	0.6	0.17	0.26	0.3	**0.33**	**0.31**	0.6	0.17	0.26
99	0.6	0.17	0.26	0.4	0.25	0.31	0.4	0.25	**0.31**
corel_10k.csv	**1**	0.1	0.18	0.5	0.2	0.29	0.7	0.14	0.23



(a) Precision and Recall



(b) F-Score

Figure 10. (a) Shows the precision and recall rate of Corel-10K dataset, (b) shows F-score with AlexNet, GoogleNet, and ResNet

Figure 10 presents the precision and recall rates achieved for Corel-10K dataset by employing different CNNs, including AlexNet, GoogleNet, and ResNet. It indicates that the CNN models exhibit highest average precision of 1.0 in most of the categories, highest recall rate 0.33 in some categories and highest f-measure score.

4. Conclusions

In conclusion, this research article presents a creative approach for image retrieval, which centers on precisely identifying shapes, objects, textures, and spatial color details. Our approach demonstrates significant capabilities in feature detection, leveraging both local and global characteristics to achieve accurate image classification. By integrating ResNet, GoogleNet, and AlexNet CNN architectures, we achieve outstanding results in tackling the challenges of image retrieval. The study evaluates performance using the Corel-10K, Cifar-10, and Tropical fruits datasets, consolidating numerous stages into a single process to improve computational effectiveness. Incorporating a scale-space pyramid within the algorithm facilitates key point identification across various scales and robust binary descriptor extraction. This aspect enhances the algorithm's adaptability to changes in lighting conditions and viewpoints, resulting in more precise object detection. Furthermore, color feature sets are computed and merged with the feature set generated by CNN sin a later stage. Lastly, the Bag-of-Words model is utilized for the retrieval of suitable images, showcasing high accuracy and precision in our methodology's outcomes.

References

1. R. Johnson and T. Zhang, "Semi-supervised convolutional neural networks for text categorization via region embedding," Adv. Neural Inf. Process. Syst., vol. 2015-January, pp. 919–927, 2015.
2. W. Zhou, H. Li, and Q. Tian, "Recent Advance in Content-based Image Retrieval: A Literature Survey," pp. 1–22, 2017, [Online]. Available: http://arxiv.org/abs/1706.06064
3. Y. Kim, "Convolutional neural networks for sentence classification," EMNLP 2014 - 2014 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf., pp. 1746–1751, 2014, doi: 10.3115/v1/d14-1181.
4. S. Hamad, A. Iqbal, S. Naz, N. ul, M. Imran, and B. AlHaqbani, "Content-Based Image Retrieval Using Texture Color Shape and Region," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 1, 2016, doi: 10.14569/ijacsa.2016.070156.
5. S. R. Dubey, S. K. Singh, and R. K. Singh, "Local neighbourhood-based robust colour occurrence descriptor for colour image retrieval," IET Image Process., vol. 9, no. 7, pp. 578–586, 2015, doi: 10.1049/iet-ipr.2014.0769.
6. J. M. Guo, H. Prasetyo, and J. H. Chen, "Content-based image retrieval using error diffusion block truncation coding features," IEEE Trans. Circuits Syst. Video Technol., vol. 25, no. 3, pp. 466–481, 2015, doi: 10.1109/TCSVT.2014.2358011.
7. M. Verma and B. Raman, "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," Multimed. Tools Appl., vol. 77, no. 10, pp. 11843–11866, 2018, doi: 10.1007/s11042-017-4834-3.
8. T. G. Subash Kumar and V. Nagarajan, "Local curve pattern for content-based image retrieval," Pattern Anal. Appl., vol. 22, no. 3, pp. 1233–1242, 2019, doi: 10.1007/s10044-018-0724-1.
9. P. S. Hiremath and J. Pujari, "Content based image retrieval using color, texture and shape features," Proc. 15th Int. Conf. Adv. Comput. Commun. ADCOM 2007, no. August, pp. 780–784, 2007, doi: 10.1109/adcom.2007.21.
10. K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," Inf. Fusion, vol. 51, no. November 2018, pp. 76–99, 2019, doi: 10.1016/j.inffus.2018.11.004.
11. K. T. Ahmed, H. Afzal, M. R. Mufti, A. Mehmood, and G. S. Choi, "Deep Image Sensing and Retrieval Using Suppression, Scale Spacing and Division, Interpolation and Spatial Color Coordinates with Bag of Words for Large and Complex Datasets," IEEE Access, vol. 8, pp. 90351–90379, 2020, doi: 10.1109/ACCESS.2020.2993721.
12. K. Kanwal, K. T. Ahmad, R. Khan, N. Alhusaini, and L. Jing, "Deep learning using isotroping, laplacing, eigenvalues interpolative binding, and convolved determinants with normed mapping for large-scale image retrieval," Sensors (Switzerland), vol. 21, no. 4, pp. 1–39, 2021, doi: 10.3390/s21041139.
13. K. T. Ahmed, S. Jaffar, M. G. Hussain, S. Fareed, A. Mehmood, and G. S. Choi, "Maximum Response Deep Learning Using Markov, Retinal Primitive Patch Binding with GoogLeNet VGG-19 for Large Image Retrieval," IEEE Access, vol. 9, pp. 41934–41957, 2021, doi: 10.1109/ACCESS.2021.3063545.
14. K. Tehseen, A. Muhammad, and A. Iqbal, "Region and texture based effective image extraction," Cluster Comput., vol. 21, no. 1, pp. 493–502, 2018, doi: 10.1007/s10586-017-0915-3.
15. K. T. Ahmed, S. A. H. Naqvi, A. Rehman, and T. Saba, "Convolution, approximation and spatial information based object and color signatures for content based image retrieval," 2019 Int. Conf. Comput. Inf. Sci. ICCIS 2019, 2019, doi: 10.1109/ICCISci.2019.8716437.
16. K. T. Ahmed, A. Irtaza, and M. A. Iqbal, "Fusion of local and global features for effective image extraction," Appl. Intell., vol. 47, no. 2, pp. 526–543, 2017, doi: 10.1007/s10489-017-0916-1.
17. K. Kanwal, K. T. Ahmad, R. Khan, A. T. Abbasi, and J. Li, "Deep learning using symmetry, FAST scores, shape-based filtering and spatial mapping integrated with CNN for large scale image retrieval," Symmetry (Basel)., vol. 12, no. 4, p. 612, 2020, doi: 10.3390/SYM12040612.
18. Krizhevsky, "CIFAR-10 dataset." 2010. [Online]. Available: https://www.cs.toronto.edu/~kriz/index.htmlhttps://www.google.com/search?q=cifar+10+dataset+images&sca_esv=575734426&tbm=isch&sxsrf=AM9HkKlY5uaiNaTBfmVih-skx0Pa53Bjkw:1698051961231&source=lnms&sa=X&ved=2ahUKEwio_dW_6IuCAxVfh_0HHaQKMWoQ_AUoAXoECAIQAw&biw=1821&bih=825&dpr=0.75#imgrc=aleaEOAxsXCgBM
19. "Corel 10K dataset." https://www.google.com/search?q=corel-10%2C000+dataset+images&tbm=isch&ved=2ahUKEwjBw8XQ6YuCAxU5pCcCHYwUDkEQ2-cCegQIABAA&oq=corel-10%2C000+dataset+images&gs_lcp=CgNpbWcQAzoECCMQJ1CiC1jkKGD_MGgAcAB4AIABxQaIAZwmkgELMi03LjEuMy4xLjGYAQCgAQGqAQtnd3Mtd2l6LWltZ8ABAQ&sclient=img&ei=qTg2ZYHLArnInsEPjKm4iAQ&bih=825&biw=1821#imgrc=zXrGIbkQrwYcTM
20. "Fruits-and-Vegetables-Dataset-6." https://www.google.com/search?q=tropical-fuits+dataset+images&tbm=isch&ved=2ahUKEwiX267D6IuCAxX1ticCHZQwGlcQ2-cCegQIABAA&oq=tropicalfuits+dataset+images&gs_lcp=CgNpbWcQAzoECCMQJzoHCAAQGBCABFDjB1jYNWCXOWgCcAB4A4ABByQeIAeJBkgELMi04LjUuMi4xLjSYAQC

gAQGqAQtnd3Mtd2l6LWltZ8ABAQ&sclient=img&ei=gDc2ZZebO_XtnsEPlOHouAU&bih=825&biw=1821#imgrc=3rTAcaF3FoioOM

21. Phalguni Singh Ngangbam. (2023). Implementation of Content-Based Image Retrieval Using Artificial Neural Networks. Engineering Proceedings.
22. Jaber, M.M., & Yildirim, M., & Cinar, A. (2023). Deep Learning for Content-Based Image Retrieval in FHE Algorithms. Applied Sciences.
23. Zhang, H., & Liu, Y., & Wang, J. (2023). RetCCL: Clustering-guided Contrastive Learning for Whole-Slide Image Retrieval. Journal of Applied Sciences.
24. Li, X., & Zhao, Y., & Wang, Y. (2023). Deep Hashing Image Retrieval Based on Hybrid Neural Network and Transformer. IEEE Transactions on Neural Networks and Learning Systems.
25. Doe, J., & Smith, A., & Lee, C. (2023). Deep Learning-Based Retrieval Algorithms for Recommendation Systems. IEEE Xplore.
26. Chen, H., & Wang, J., & Zhang, Y. (2023). Spherical Centralized Quantization for Fast Image Retrieval. IEEE Transactions on Image Processing.
27. Qureshi, S., & Rehman, S., & Khan, I. (2023). A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion. Applied Sciences.
28. Abbas, F., Iftikhar, A., Riaz, A., Humayon, M., & Khan, M. F. (2024). Use of Big Data in IoT-Enabled Robotics Manufacturing for Process Optimization. Journal of Computing & Biomedical Informatics, 7(01), 239-248.
29. Batool, S., Abid, M. K., Salahuddin, M. A., Aziz, Y., Naeem, A., & Aslam, N. (2024). Integrating IoT and Machine Learning to Provide Intelligent Security in Smart Homes. Journal of Computing & Biomedical Informatics, 7(01), 224-238.
30. Hussain, S.K., Ramay, S.A., Shaheer, H., Abbas T., Mushtaq M.A., Paracha, S., & Saeed, N. (2024). Automated Classification of Ophthalmic Disorders Using Color Fundus Images, Volume: 12, No: 4, pp. 1344-1348 DOI:10.53555/ks.v12i4.3153
31. Ammar Ahmad Khan , Muhammad Arslan , Ashir Tanzil , Rizwan Abid Bhatty , Muhammad Asad Ullah Khalid , Ali Haider Khan. (2024). Classification Of Colon Cancer Using Deep Learning Techniques On Histopathological Images. Migration Letters, 21(S11), 449–463.