Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 02
2024

Research Article
https://doi.org/10.56979/702/2024

Embedded Descriptor Carriers Computation using Multi-Layer Neural Networks on Large Datasets

Khawaja Tehseen Ahmed[1], Sidra Sarwar[1*], Aiza Shabbir[1], Syed Burhan ud Din Tahir[2], Naila Sattar[3], Nosheen Saeed[3], Adeeba Rashid Khan[3], and Sobia Sarfraz[3]

[1]Department of Computer Science, Bahauddin Zakariya University, Multan, 60800, Pakistan.
[2]Department of Computer Science, Air University, Multan, Pakistan.
[3]Department of Computer Science, NCBA&E University, Gulghasht Colony, Multan, Pakistan.
[*]Corresponding Author: Sidra Sarwar. Email: sid.sarwer@gmail.com

Received: March 22, 2024 **Accepted:** July 12, 2024 **Published:** September 01, 2024

Abstract: Intelligent and effective visual extraction due to extensive data sets is unavoidable need included in today. Raw image labels must correspond to visual characteristics in order to extract images based on content based image retrieval (CBIR). CBIR is extensive method progressively applied on retrieval methods. CNN major job is to retrieve authentic and useful images. Various methods have been employed to enhance the effectiveness and reliability of image exploration, such as filename-based searches and image tagging. However, these techniques have not proven successful in real-world applications. To effectively categorize images and function as a filter, the feature vector must include comprehensive visual information. This information should encompass elements such as color, shape, objects, and different types of spatial data. By incorporating these details, the feature vector can more accurately define the image's category and improve the overall efficiency of image exploration. The proposed method excels at detecting, describing, recognizing, and correlating image signatures that accurately reflect the true content of an image. It achieves this by categorizing semantic groupings of nearly identical images. This technique is particularly effective during image retrieval and feature detection processes. The provided methodology details a convolutional neural network (CNN) based method for colorizing grayscale images. The approach begins with defining the area of interest and potentially converting the image to grayscale. Pixel intensities are compared, and patterns within the image are identified using techniques like concentric, retinal, or log-polar methods. Selective pixel sampling, differentiation to analyze neighboring pixels, and smoothing to reduce noise are all employed. Convolution and pooling operations further refine the data. Activation functions, like ReLU or SoftMax, are then applied. Finally, fully connected layers likely within a neural network come into play. The later stages involve sample collection, redundancy measures, distance calculations, and potentially techniques like Bag-of-Words (BoW) and K-Nearest Neighbors (KNN) for image classification. An FV is aggregated, and Bow, KNN, and results are generated. The presented method is capable of identifying, characterizing and correlating images signatures that precisely represent the actual content of a picture by categorizing semantically grouped, nearly identical images. The proposed technique occur during image retrieval and feature detection. Extensive experimentations are conducted on highly recognized datasets such as ALOT-250 and 17-Flowers with mismashed of RestNet-50, Inception and VGG-19. Remarkable results indicated that the presented technique demonstrates significant precision rate and recall rate for huge image groups of challenging datasets.

Keywords: Content Based Image Retrieval; Convolutional Neural Network; Spatial Data; Feature Recognition; Feature Vector.

1. Introduction

The rapid advancements in computers, information technology, and scientific fields have caused a considerable rise in production of altered and computer images. This increase has consequently introduced numerous substantial problems in image retrieval. In the realm of computer vision, various methods are employed to deal with image indexing, including examination of color, analysis of texture, segmented attributes, sampling of keywords, and detection of interest points. To address the difficulties encountered in image retrieval, CBIR systems leverage image characteristics and data. In CBIR, finding pictures based on their shapes is tough because we don't have good ways to compare shapes visually yet. Data on the various shapes can be drawn as lines and curve fashion or with mathematical equations to describe them [1].

It turns out that in order to get meaningful features from an image when considering a small part of the image the pixels in the immediate neighborhood are critically important. Feature enhancement enhances the details of values near the center pixel to focus more on smaller values adjacent to the center and combines it in a way to give more importance to images closer to the center pixel. This can make the description of each part more accurate. The use of percentages is therefore useful [2].

Individuals learning CBIR systems interpretation proficiency is mainly dependent on the attribution among features and similarity examination. One of the primary problems encountered in the contemporary CBIR research is the so-called "semantic gap", which consists in the fact that the low-level visual information acquired by the computerized method is not sufficient for a person to understand a depicted scene on a higher level of semantics [3].

In CBIR the speed and accuracy of image classification is paramount therefore there is a need for systems that can meet this supply. Feature retrieval is one of the most significant steps in CBIR because like the suggestions of the visual recognition task are also dependent upon the feature retrieval system [4].

The regions upon which shapes are identified improve our understanding of objects. Detecting corners and edges through pixel intensities enables the creation of highly effective and resilient detectors and descriptors [5].

Deep learning features are really important for finding images, especially when dealing with lots of them. They contain details like texture. This directly affects how accurate our searches [6].

Convolutional Neural Networks (CNNs) are extensively employed deep learning models, particularly for processing complex data such as images and videos. CNNs excel in handling multidimensional input data that is locally correlated, overcoming the scalability challenges faced by traditional Deep Neural Networks (DNNs) [7].

Features are extracted to represent an image. Visual features are global and local features. Global features fail to encapsulate all image characteristics. Conversely, local features effectively mitigate the semantic gap [8].

To evaluate correspondence among images, capable to utilize processes to second broadly acknowledged category, which depends upon local points (key points) or another descriptors which capture specific visual elements [9].

Color histogram descriptors prioritize color distribution over spatial arrangement, neglecting spatial information. In contrast, the color coherence vector (CCV) evaluates homogeneity of pixel arrangement in an image. For instance, if red pixels cluster together within sizable regions, the CCV assigns high coherence to red; conversely, if red pixels are dispersed, coherence diminishes [10].

CNN serves as an expressive local descriptor, designed to suppress image content effects and extract diverse residual features for splicing detection. The CNN's first layer is initialized with high-pass filters and fine-tuned to retain these properties, enhancing generalization with contrastive and cross-entropy loss. A feature fusion strategy, block pooling, extracts discriminative features for SVM-based detection, and a fully connected conditional random field (CRF) aids in localization [11].

Deep learning techniques have grown rapidly, especially in computational biology and cognitive neuroscience. Concurrently, the need to explore and manipulate deep learning models has increased. Multilayer approach outperforms single-layer methods in capturing CNN complexity and enabling manipulations [12].

The Extreme Learning Machine (ELM) is a single-hidden-layer feedforward learning algorithm widely used in regression and classification across various fields. It assigns random weights and biases in the hidden layer, using the Moore–Penrose inverse matrix in regularized least squares to compute output layer weights. ELM is known for its training speed, generalization ability, and robustness, but it struggles with highly nonlinear problems. To address this, non-iterative multilayer learning models like the Multilayer Extreme Learning Machine (which uses an unsupervised extreme learning Autoencoder for feature mapping) have been developed [13].

The methodology outlined in this study introduces a convolutional neural network (CNN) based strategy for colorizing grayscale images. Initially, the input image undergoes conversion to grayscale. Subsequently, it undergoes comparison with pairs of pixel intensities, followed by a sequence of equally spaced circular patterning and concentric circle formation. This is succeeded by retinal sampling, coefficient formation, differentiation, smoothing, and convolution, with spatial padding and max pooling applied subsequently. A color feature vector is then generated, and activation is carried out. Mask sizing and kernel induction, along with exponential adjustment, are executed, leading to the creation of overlapped fields and the application of standard deviation. Samples are collected and log-polar patterns utilized, with overlapping and redundancy implemented, followed by difference thresholding. Massive pairing and spatial distance computation ensue, culminating in the aggregation of an FV and the generation of results through Bow, KNN, and other methods. Extensive experimentation is conducted on well-known datasets such as ALOT-250 and 17-Flowers, leveraging ResNet-50, Inception, and VGG-19. ResNet, an abbreviation for residual network, denotes a critical layer set frequently employed to enhance neural network resolution. VGG-19, an extension of the VGG model, comprises 19 layers primarily designed for image classification. Remarkable outcomes highlight the presented technique's substantial precision and recall rates, particularly for large image datasets containing diverse challenges.

The formulation of this paper is summarized as: Section 2 offers extensive examination of the literature concerning convolutional neural networks, deep learning, and content-based image retrieval. Section 3 contains comprehensive explanation about the new approach. Section 4 outlines about experimental outcomes, while Section 5 serves as this paper's conclusion.

2. Related Work

With the rising number of digital images nowadays, finding a specific picture in a large database can be quite challenging. To address this issue, new and highly effective image retrieval techniques have been developed. In past, CTDCIRS employs adaptive primary color and pattern co-occurrence matrix, and scan pattern difference for retrieval. It divides images into partitions and represents texture using MCM and DBPSP. Integration of dominant color, MCM, and DBPSP enhances retrieval efficiency, supported by experimental evidence [14]. Researchers proposed an experimental architecture for a CBIR laboratory system aimed at analyzing visual features, aiming to link the divide among automatic initial level extraction and human-level expression. The software system "Art Painting Image Color Aesthetics and Semantics" (APICAS) enables cataloging and retrieving in painting collections based on content attributes. [15].

The 3D color histogram and Gabor filter describe image properties effectively. Proposed method [16] combines color coherence vector and wavelets for improved retrieval. Feature selection, using discrimination and Genetic Algorithm, refines features. Researcher implemented a retina-inspired key point descriptor to enhanced performance [17].

Structure Elements' Descriptor (SED) for image texture description [18] to effectively extract characteristics related to appearance, Histogram of Image Structure Elements (SEH) is calculated using SED in the HSV color space with 72 bins. SEH combines analytical and architectural methods for pattern depiction, enhancing to capture of spatial correlation. CBIR systems offer a means to extract visuals with comparable semantics from large databases. The academic community is in search of more effective CBIR techniques for time-sensitive applications.

This paper [19] introduced a NN for CBIR, which incorporates an effective feature extraction technique. This technique utilizes intricate pattern assessment via wavelet decompositions and Gabor filters characteristics applied to visual representation. Comparative analysis showcases the superior efficiency of the presented approach. Proposed [20] introduced image database creation, visual feature extraction for retrieval. Validation via color histogram and Euclidean distance on database. System includes design, extraction, distance techniques. Achieves 85% accuracy over 20 iterations. The researchers introduced a proficient description of color-based image features for CBIR.

The descriptor [21] achieves inherent rotation invariance by quantifying color occurrences in local pixel neighborhoods. There are 64 shades within the RGB color space for streamlined representation. Color occurrences are extracted and depicted in dual form to generate binary arrangements based upon local color occurrences, ensuring the effective capture of local color information. A novel approach to image extraction combines form, textural and coloration elements within CBIR. Simulation results demonstrate good precision, particularly for clear and distinct targets in queries, showcasing the method's efficiency. Notably, the texture feature, referred to as CLCM, outperforms GLCM in retrieval. However, there's room for substantial improvement, especially for complex query scenes where low-level visual features may be insufficient. Utilizing high-level visual features like semantic features could further enhance retrieval performance [22].

This paper introduces an effective scheme for enhancing image quality in video surveillance, enhancing face detection and introducing a content-based image retrieval method. The integrated enhancement method combines contrast enhancement and color balancing, showing significant improvements in face detection. The retrieval approach prioritizes results using fuzzy heuristics and employs well-known algorithms for color, texture, and shape analysis. Evaluation metrics demonstrate superior performance compared to existing methods, with future work focusing on real-world data integration and processing speed considerations [23].

Image signatures are derived through the aggregation of interest points across different representation levels, initially gathering shape features by grouping connected pixels based on binary brightness thresholds. The HOG is utilized to characterize attributes for identified points of interest within optimal stability regions. Descriptors are then combined with rotation-invariant texture attributes obtained through uniform patterns following the application of our suggested reordering algorithm. Experimentation on Caltech-101, Caltech-256 and Corel-100, and datasets illustrates the superiority of our technique over existing methods across various image categories. Results indicate that the integration of local and global features enhances the method's capability in foreground and background object retrieval, with feature description employing a sliding window approach, thereby strengthening its robustness in object recognition [24].

Arabic handwritten script recognition, challenging due to diverse styles and large datasets, benefits from effective techniques. BDLF-MLP, using Block Density and Location Feature with MLP, achieves 97.26% accuracy by extracting pixel density and location from letter images, surpassing other algorithms [25].

Convolutional neural networks (CNNs) exploit grid-structures and spatial dependencies in images to detect adjacencies, colors, and patterns. This research integrates GoogLeNet, VGG-19, and ResNet50 with Eigenvalues and convolutional Laplacian scaled object features, using mapped colored channels for high image retrieval rates across diverse benchmarks. The approach enhances deep learning fusion and descriptor creation efficiency, showing remarkable performance on datasets like ALOT, Corel-1000, CIFAR-10, CIFAR-100, Oxford Buildings, and others, compared to state-of-the-art methods, demonstrating significant accuracies across various image types [26]

This study enhances image retrieval by detecting interest points and utilizing their features for content-based analysis, including shape, texture, and color attributes. It improves feature detection with techniques such as non-maximum suppression, edge and corner detection, and symmetric sampling. Evaluation on benchmarks like Corel-1000, ImageNet, and Caltech-101 shows superior performance, comparing favorably with RGBLBP, LBP, SURF, SIFT, DoG, HoG, and MSER in accuracy and retrieval rates [27].

3. Methodology

In figure 1, CNN-based approach for colorizing grayscale images is presented. First, the image is converted to grayscale and compared with pixel intensity pairs. Circular patterns and concentric circles are formed, followed by retinal sampling, coefficient formation, differentiation, smoothing, and convolution with spatial padding and max pooling. A color feature vector is generated, activated, and processed through mask sizing and kernel induction. Exponential adjustment creates overlapped fields, and standard deviation is applied. Samples are collected using log-polar patterns, with overlapping and redundancy, followed by difference thresholding. Massive pairing and spatial distance computation culminate in feature vector aggregation. Results are generated through BoW, KNN, and other methods. Extensive experimentation on datasets like ALOT-250 and 17-Flowers using ResNet-50, Inception, and VGG-19 demonstrates the technique's high precision and recall rates.

3.1. Smoothing technique to reduce the impact of noise.

Pixel intensity analysis uses pairwise comparisons to create binary sequences, inspired by human retina functions. A circular pattern grid compares pixel densities, with Fast Retina Key points identifying key points in Retina Sampling mode to exploit density differences. Noise reduction normalizes each point, with circular patterns in BRISK made of evenly spaced dots in concentric circles for computational efficiency. Gray values are determined by these circles, with object orientation inferred from local gradients. Retinal sampling mimics the human retina, increasing density toward the center to capture detailed features, with circles representing Gaussian kernel standard deviation.

3.2. Overlapping

Our research introduces adaptive kernel sizes within patches, inspired by the retinal model's varying receptive field sizes, enhancing visual information extraction. Exponential kernel size variations improve flexibility and accuracy, fostering adaptability and discriminative capabilities. Overlapping receptive fields capture spatial dependencies and contextual information for accurate scene interpretation. Gaussian kernels, controlled by a log-polar retina grid, provide smooth weighting, optimizing performance in computer vision tasks. Aligning kernel sizes with the retinal sampling grid boosts performance and discriminative power during pattern extraction.

3.3. Utilize redundancy

In our research, we enhanced our method by introducing redundancy through the integration of overlapping receptive fields. This enhancement strengthened the discriminative capacity of our method. Through redundant representation of information across these domains, we have enhanced the robustness and reliability of the extracted attributes. This redundancy serves as a mechanism for robustness, facilitating the integration of information and agreement among various points. Even in scenarios where certain points are affected by noise or variations. This strategy, driven by redundancy, mitigates the impact of noise and image variations, leading to more dependable and identifying feature descriptors.

Let's consider the intensity degrees Ij evaluated at the perceptive fields P,Q, and R in which :

$$I_P > I_Q,\ I_Q > I_R, \text{ and } I_P > I_R \qquad (1)$$

In scenarios where the fields are not overlapping, the ultimate test IP >IR

3.4. Implement filtering for variances and comprehensive matching:

Our descriptor in binary form, labeled as E, is formed through implementing a filtering to the comparison of permissible field combinations along with Gaussian filters. Consequently, E comprises a series of Gaussian differences (DoG), constructing a binary pattern.

$$E = {}^{X} 2^{b} A\ (R_b), \qquad (2)$$

$0 \leq b < P$ in which R_b is a pair of perceptive fields, P is the optimal dimensions of the descriptor, and

$$A(R_L) = \{1 \quad if\ (I(R_b^{r1}) - b\ 0 \quad otherwise, I\left(R_{\square}^{r2}\right) > 0 \qquad (3)$$

With $I(R_b^{r1})$is stabilized intensity of the initial perceptive area in the matching set. R_b.

While a several perceptive area can generate numerous possible pairs, to identify that not all of these pairs are crucial for efficiently defining an image.

3.5. CNN

Evaluation of modified deep neural network performance and accuracy entails assessing three architectures VGG-19, GoogleNet and ResNet implemented on image data sets. GoogleNet, referred to as Inception-v, features 316 layers and 350 connections. Utilizing ResNet-50, VGG19, and GoogleNet layers, the proposed approach aggregates feature vectors extracted by CNNs and heuristic key points, inputting them into a BoW approach for deep image fetching. Grayscale conversion of request images enhances pixel intensity analysis, crucial for visual content analysis, with color attributes aiding image distinction and recognition. Bag-of-Words model, based on feature extraction, ranks images by shared features with query images, albeit without considering spatial information. The K-nearest neighbor's algorithm, employed for classification, determines class based on the most frequent occurrence among k-closest neighbors of unknown data points, complementing Bag-of-Words outcomes represented by feature histograms.

4. Experimentation

The IR system's efficiency and precision are evaluated by employing diverse image databases containing a range of elements, including object information, spatial color, complexity, and the general application of CBIR. To enhance efficiency and accuracy, experiments are performed across diverse datasets, utilizing CBIR to analyze visual attributes like shape, texture, color and object information. Two datasets covering a broad spectrum of categories are examined, sourced from various origins to observe a wide array of objects. The numerical representation of results for each database's categories reflects differing levels of accuracy and effectiveness, attributed to multiple classifications and diverse object information within each database. Established benchmarks like 17-Flowers and A LOT-250 are employed in conducting experiments.

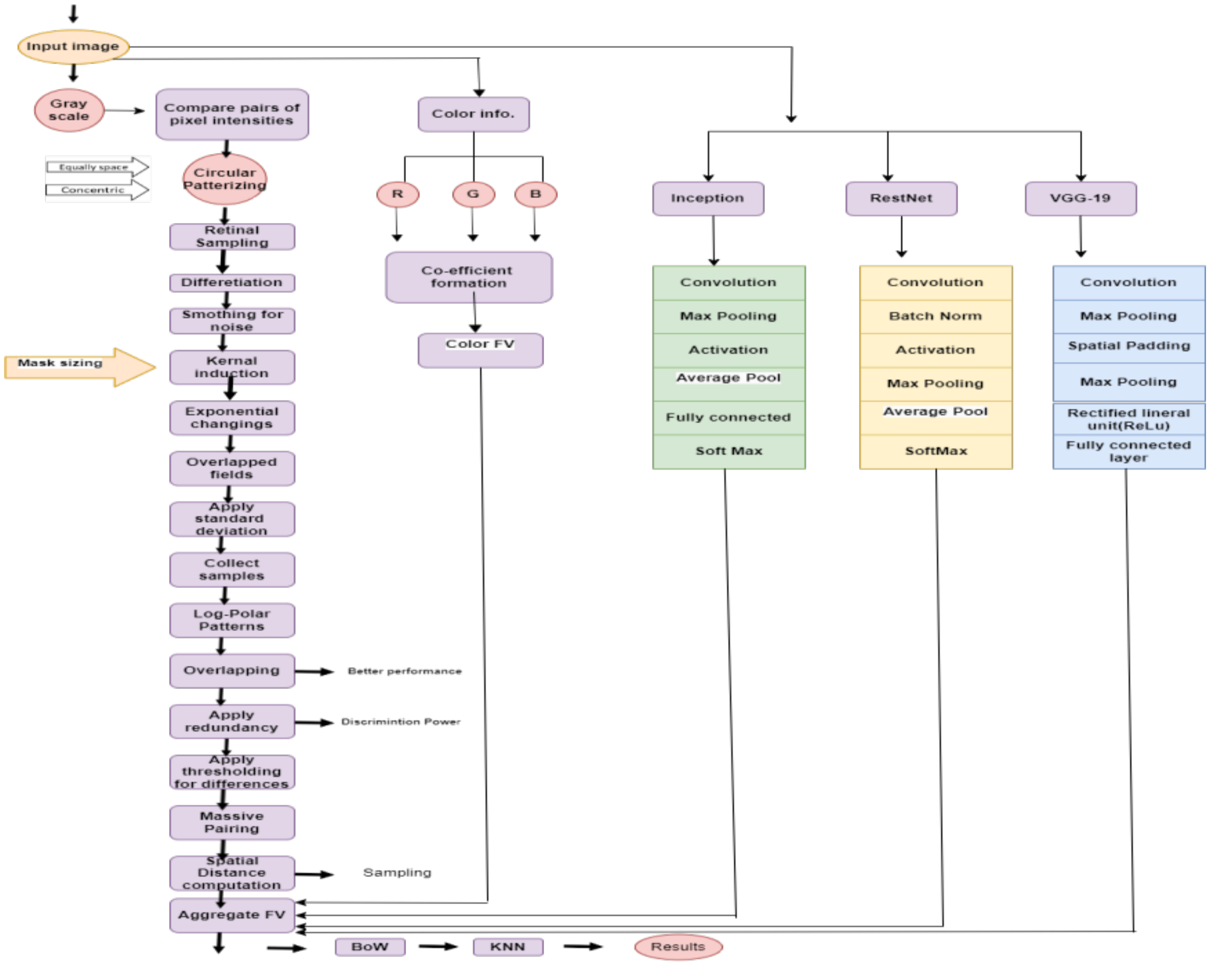


Figure 1. Proposed Methodology

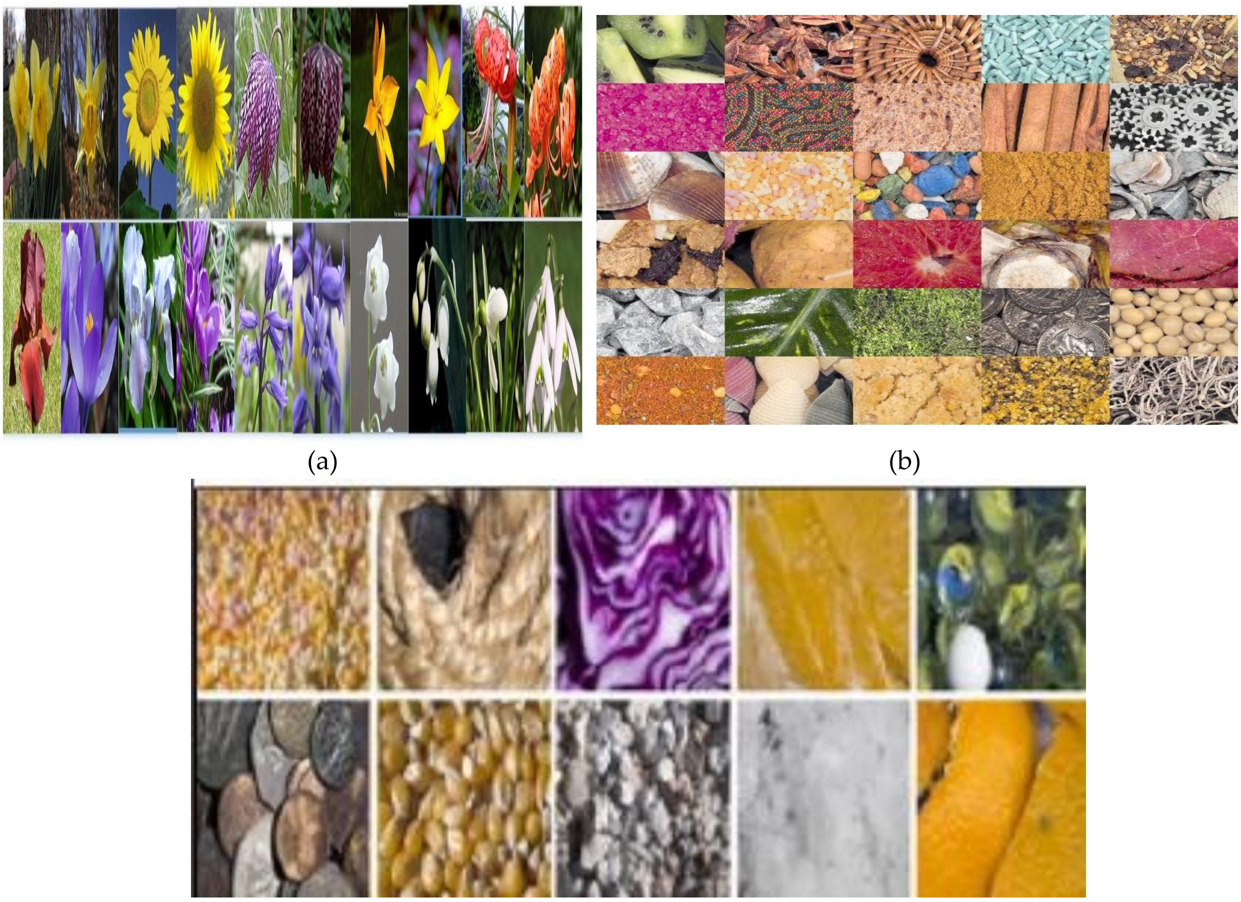
4.1. Datasets

4.1.1. 17-Flowers

Dataset 17 flowers comprises 17 classes, containing images sized 17 by 80 pixels, depicting various flowers. System performance is assessed using metrics like average accuracy and recall with this dataset. Image groups within include hyacinth, primrose, ranunculus, and others, each category comprising 80 images, totaling 1360. Generally attributes like shape and color are examined across categories. The presented method demonstrates promising fitness scores for the majority of flower groups, depicted in Figure 2(a).

4.1.2. ALOT-250

The ALOT dataset comprises 250 color images showcasing diverse and irregular textures, gathered for research purposes, with each category containing 100 images at a resolution of 384 * 235 pixels, totaling 2500 images. From this repository, 10 categories were selected for experimentation, including rope, fruit sprinkles, toy marble, and others, as showcased in figure 2(b). Evaluation using this dataset, covering 250 species, along with various other datasets such as pasta, sand, fruit, and more for assessment. The presented projection method efficiently clusters texture images according to similarities in foreground and background objects, categorizing them into general categories for precise description, achieving notable results with up to 80% accuracy even in complex categories as indicated in Figure 2(c).



(a) (b)

(c)

Figure 2. (a) Shows 17-flowers data set images, (b) and (c) shows ALOT-250 data set images

4.2. Evaluation of Precision, Retrieval rate and F1 score

4.2.1. Precision

Precision is computed based on the predicted positive values.

$$\textbf{Precion} = \frac{Ew(n)}{Eu(m)} \tag{4}$$

4.2.2. Recall

Recall is derived from the true positive ratio.

$$\textbf{Recall} = \frac{Ew(m)}{Eo} \tag{5}$$

4.2.3. F1 score

The F1 score is calculated by obtaining harmonic mean of retreival rate (t) and accuracy rate (s), as per the formula. Equation (6).

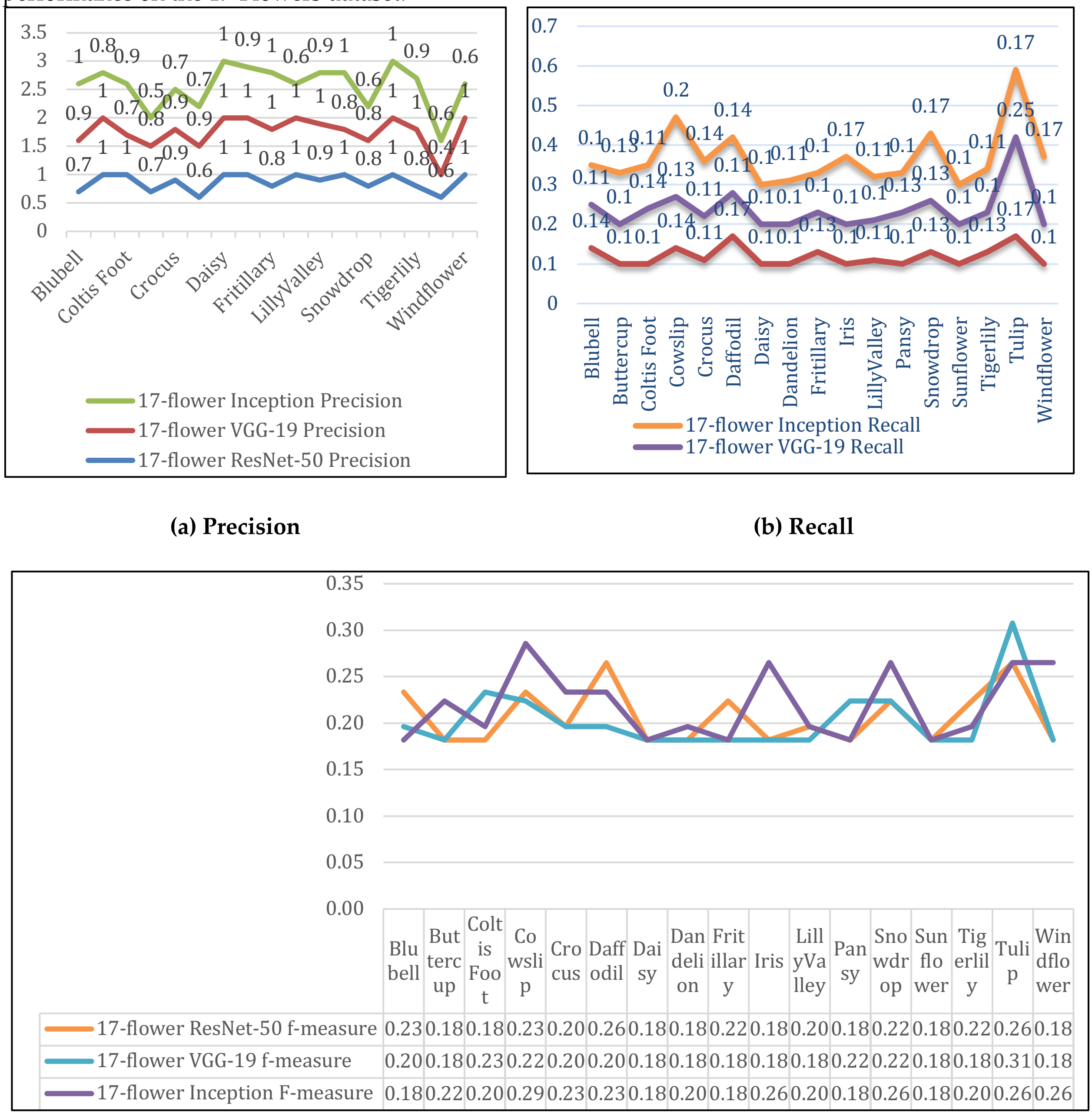
$$f = \frac{2\times s\times t}{s+t} \tag{6}$$

4.3. Observations and Outcomes of the Datasets

The suggested method employs color images sourced from benchmark datasets as its input, which undergo processing by a convolutional neural network (CNN). Surprisingly, the system initially converts these color images to grayscale. The efficacy and precision of the image retrieval system are notably impacted by selecting of suitable image datasets. We conducted experiments on three separate benchmark datasets—17-Flowers and ALOT-250 to measure the accuracy.

4.3.1 Evaluations on 17-Flowers Dataset

Our proposed method was evaluated using the extensive 17-Flowers dataset, which features various semantic groups like lily of the valley, yellow primrose, crocus, dandelion, snowdrop, hyacinth, tulip, sunflower pansy, ranunculus, hazel grouse, daisy, iris, coltsfoot and narcissus. Across most 17-Flowers categories, our method demonstrates significantly improved average precision (AP). Employing CNN features, our approach achieves accurate image classification, covering diverse semantic categories such as lily of the valley, yellow primrose, crocus, dandelion, snowdrop, hyacinth, tulip, sunflower pansy, ranunculus, hazel grouse, daisy, iris, coltsfoot and narcissus. Notably, our proposed approach achieves an average precision exceeding 95% on the 17-Flowers dataset. Figure 3 visually represents our approach's performance on the 17-Flowers dataset.



(a) Precision

(b) Recall

	Blubell	Buttercup	Coltis Foot	Cowslip	Crocus	Daffodil	Daisy	Dandelion	Fritillary	Iris	LillyValley	Pansy	Snowdrop	Sunflower	Tigerlily	Tulip	Windflower
17-flower ResNet-50 f-measure	0.23	0.18	0.18	0.23	0.20	0.26	0.18	0.18	0.22	0.18	0.20	0.18	0.22	0.18	0.22	0.26	0.18
17-flower VGG-19 f-measure	0.20	0.18	0.23	0.22	0.20	0.20	0.18	0.18	0.18	0.18	0.18	0.22	0.22	0.18	0.18	0.31	0.18
17-flower Inception F-measure	0.18	0.22	0.20	0.29	0.23	0.23	0.18	0.20	0.18	0.26	0.20	0.18	0.26	0.18	0.20	0.26	0.26

(c) F score

Figure 3. (a) Shows the Average Precision (AP), (**b)** shows Recall, **c** shows f-score of 17-flowers dataset with ResNet-50, Inception, and VGG-19 architecture.

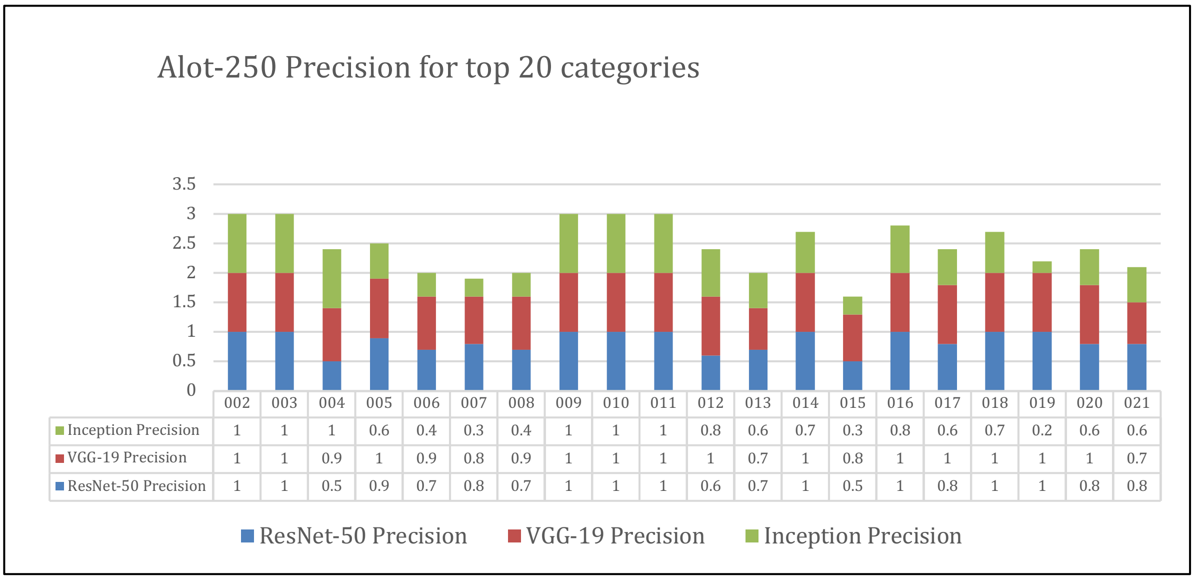
The total accuracy of the 17-flowers dataset documented as 0.90. The proposed method achieved 100% precision in various categories such as windflower, sunflower, pansy, iris, dandelion, daisy, coltis foot and buttercup. Categories include Lilly valley and crocus achieved an average precision 90% with ResNet-50. Several categories such as crocus, buttercup, Tigerlilly, pansy, Lilly valley, bluebell, sunflower and coltisfoot maintain precision rates of 70% or higher using Inception. When utilizing VGG-19 on 17-flowers, over fifty percent of the classes like crocus, cowslip, daffodils, pansy and snowdrop display accuracy rates surpassing 70%. Similarly, with Inception on 17-flowers, over other fifty percent of the classifications like tigerlilly, sunflower, iris and dandelion achieved accuracy rates exceeding 80%. Employing VGG-19 for feature extraction, the categories cowslip, tulip and snowdrop exhibit recalls of 0.13, 0.25, and 0.13, respectively, with most results falling within the 60th percentile or lower. When using ResNet-50 for feature extraction, few classifications cowslip, snowdrop and bluebell achieve recalls of 0.14, 0.13, and 0.14, correspondingly. In case of Inception, categories like tulip, iris, and daffodil have recalls of 0.25, 0.1 and 0.11. The proposed method demonstrates f-scores 0.23, 0.22 and 0.18 for coltsfoot, cowslip, lilly valley, and 0.31 for tulip when using VGG-19.The daisy category f-scores 0.18 with ResNet and Inception.

Table 1. Precision, recall and f-score values for the 17-Flowers dataset with ResNet-50, VGG-19 and Inception

17-Flowers Dataset									
	ResNet-50			**VGG-19**				**Inception**	
Category	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Windflower	1	0.1	0.18	1	0.1	0.18	0.6	0.1	0.26
Tulip	0.6	0.17	0.26	0.4	0.25	0.31	0.6	0.25	0.26
Tigerlily	0.8	0.13	0.22	1	0.1	0.18	0.9	0.1	0.2
Sunflower	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
Snowdrop	0.8	0.13	0.22	0.8	0.13	0.22	0.6	0.13	0.26
Pansy	1	0.1	0.18	0.8	0.13	0.22	1	0.13	0.18
LillyValley	0.9	0.11	0.2	1	0.1	0.18	0.9	0.1	0.2
Iris	1	0.1	0.18	1	0.1	0.18	0.6	0.1	0.26
Fritillary	0.8	0.13	0.22	1	0.1	0.18	1	0.1	0.18
Dandelion	1	0.1	0.18	1	0.1	0.18	0.9	0.1	0.2
Daisy	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
Daffodil	0.6	0.17	0.26	0.9	0.11	0.2	0.7	0.11	0.23
Crocus	0.9	0.11	0.2	0.9	0.11	0.2	0.7	0.11	0.23
Cowslip	0.7	0.14	0.23	0.8	0.13	0.22	0.5	0.13	0.29
Coltis Foot	1	0.1	0.18	0.7	0.14	0.23	0.9	0.14	0.2
Buttercup	1	0.1	0.18	1	0.1	0.18	0.8	0.1	0.22
Blubell	0.7	0.14	0.23	0.9	0.11	0.2	1	0.11	0.18

4.3.2 Experimentation on ALOT-250 Dataset

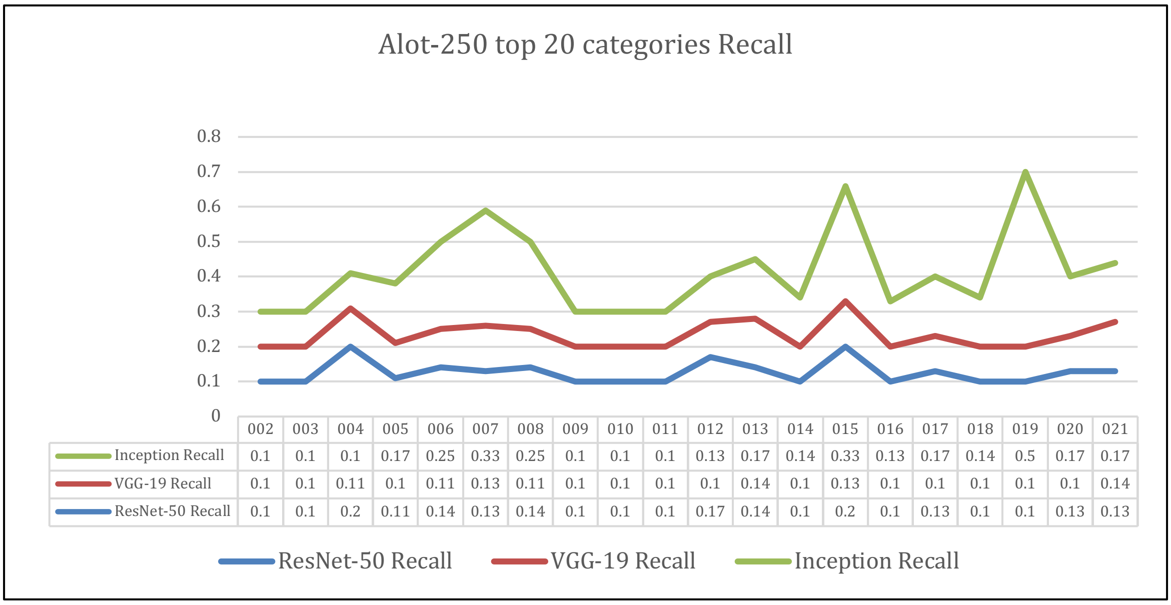
The ALOT dataset includes 250 color images with diverse textures, each class containing 100 photos at 384 x 235 pixels resolution. Ten categories were selected for experimentation, shown in Figure 4.8, with effectiveness assessed using the ALOT database comprising 250 species. Various datasets were used for evaluation, encompassing textures like pasta, sand, fruit, and more. The proposed method effectively clusters texture images based on similarities, achieving notable results in categorization and classification using CNN features. Notably, the ALOT collection demonstrates up to 80% accuracy even in complex categories.



(a)

Figure 4. (a) Describes 1-20 categories precision for ResNet-50, VGG-19 and Inception.

The graphical depiction in figure 4a showcases the top 20 categories, while table 2 presents the associated numerical data, featuring category names and precision rates. Categories such as 004, 015 achieve 50% accuracy when utilizing ResNet-50. In VGG-19 categories 013 and 021 having 70% precision. Remarkably, categories such as 002, 003, 009, 010, 011, 014, 016, and 018 achieve a 100% Average Precision (AP) when utilizing the VGG-19, ResNet-50 and Inception architecture.



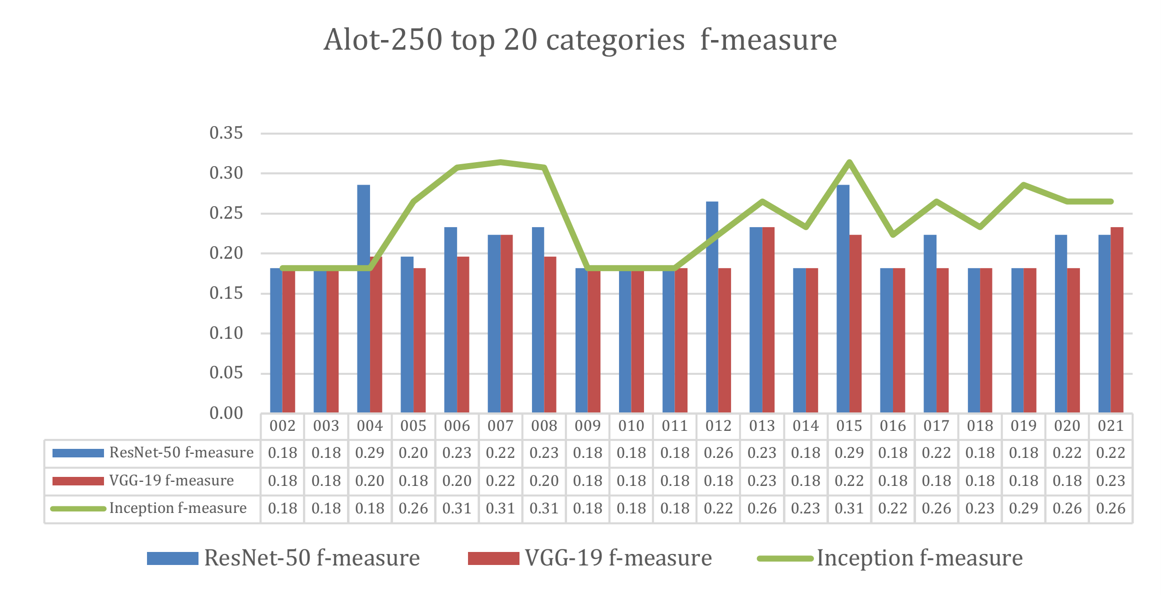
(b)

Figure 4. (b) Describes 1-20 categories Recall for ResNet-50, VGG-19 and Inception.

The graphical depiction in figure 4b showcases the top 20 categories, while table 2 presents the associated numerical data, featuring category names and recall rates. Remarkably, highest recall rate is 0.17, 0.14 and 0.33 while utilizing the VGG-19, ResNet-50 and Inception architecture. The category 012 shows 0.17 recall rate with ResNet. In VGG-19 category 007 and 015 with recall rate 0.13, 006 category recall rate 0.25 with Inception and 015 has 0.33 recall rate with Inception architecture.

The graphical depiction in figure 4(c) showcases the top 20 categories, while table 2 presents the associated numerical data, featuring category names and f-score rates. Remarkably, highest f-score rate is

0.26, 0.29 and 0.31 while utilizing the VGG-19, ResNet-50 and Inception architecture. In ResNet-50 category 004 has highest recall rate, highest f-score of category 013 and 021 is 0.23 with VGG-19. In Inception highest f-score is 0.26 for category 005, 013 and 021.



	002	003	004	005	006	007	008	009	010	011	012	013	014	015	016	017	018	019	020	021
🟦 ResNet-50 f-measure	0.18	0.18	0.29	0.20	0.23	0.22	0.23	0.18	0.18	0.18	0.26	0.23	0.18	0.29	0.18	0.22	0.18	0.18	0.22	0.22
🟥 VGG-19 f-measure	0.18	0.18	0.20	0.18	0.20	0.22	0.20	0.18	0.18	0.18	0.18	0.23	0.18	0.22	0.18	0.18	0.18	0.18	0.18	0.23
— Inception f-measure	0.18	0.18	0.18	0.26	0.31	0.31	0.31	0.18	0.18	0.18	0.22	0.26	0.23	0.31	0.22	0.26	0.23	0.29	0.26	0.26

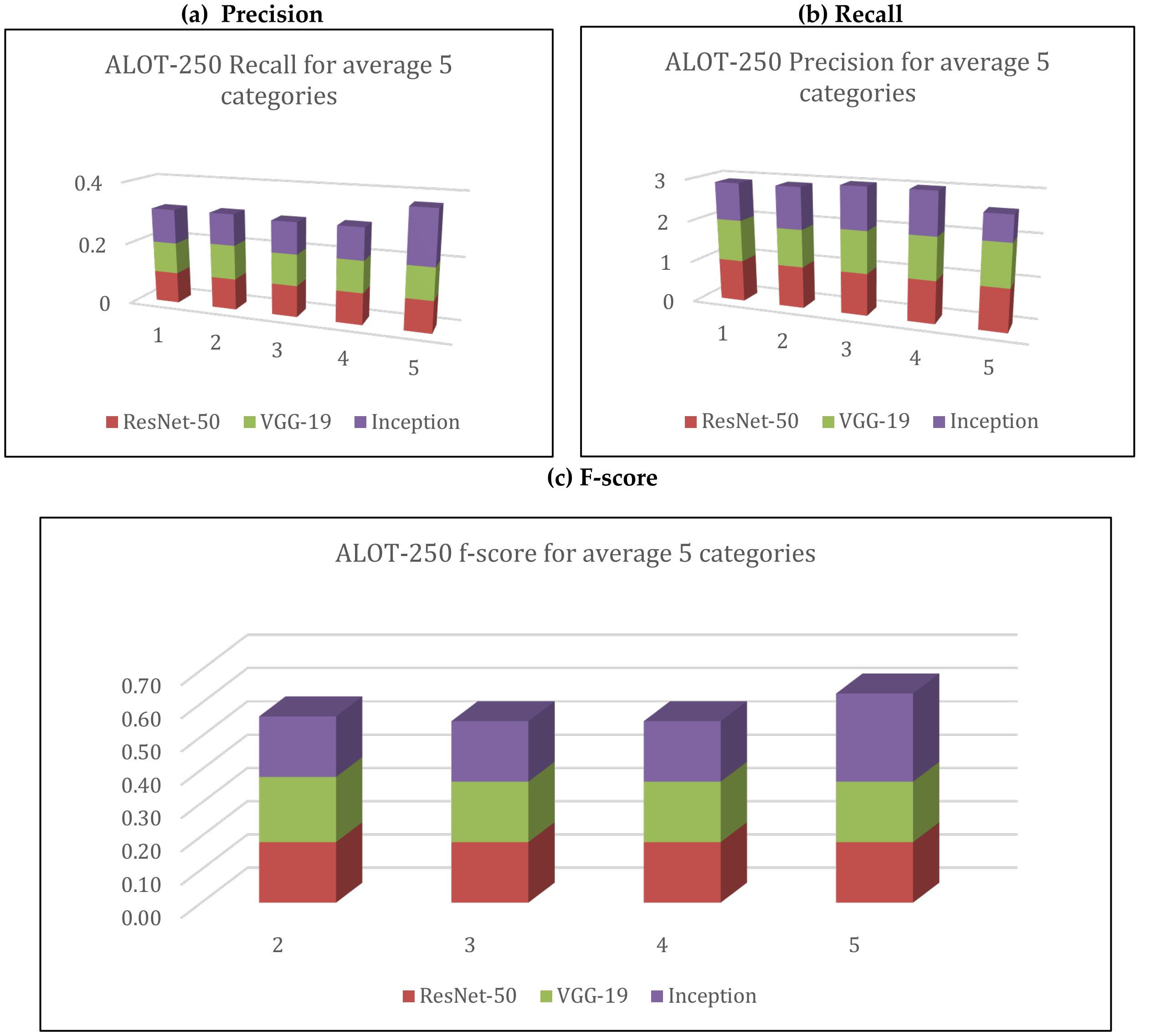
(c)

Figure 4. (c) Describes 1-20 categories f-score for ResNet-50, VGG-19 and Inception.

Table 2. Recall, precision and f-score with ResNet-50, VGG-19, and Inception of ALOT-250 dataset top 20 categories

	ALOT-250 Dataset top 20 categories								
	ResNet-50				**VGG-19**			**Inception**	
Category	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
002	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
003	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
004	0.5	0.2	0.29	0.9	0.11	0.20	1	0.1	0.18
005	0.9	0.11	0.20	1	0.1	0.18	0.6	0.17	0.26
006	0.7	0.14	0.23	0.9	0.11	0.20	0.4	0.25	0.31
007	0.8	0.13	0.22	0.8	0.13	0.22	0.3	0.33	0.31
008	0.7	0.14	0.23	0.9	0.11	0.20	0.4	0.25	0.31
009	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
010	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
011	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
012	0.6	0.17	0.26	1	0.1	0.18	0.8	0.13	0.22
013	0.7	0.14	0.23	0.7	0.14	0.23	0.6	0.17	0.26
014	1	0.1	0.18	1	0.1	0.18	0.7	0.14	0.23
015	0.5	0.2	0.29	0.8	0.13	0.22	0.3	0.33	0.31

016	1	0.1	0.18	1	0.1	0.18	0.8	0.13	0.22
017	0.8	0.13	0.22	1	0.1	0.18	0.6	0.17	0.26
018	1	0.1	0.18	1	0.1	0.18	0.7	0.14	0.23
019	1	0.1	0.18	1	0.1	0.18	0.2	0.5	0.29
020	0.8	0.13	0.22	1	0.1	0.18	0.6	0.17	0.26
021	0.8	0.13	0.22	0.7	0.14	0.23	0.6	0.17	0.26



(a) Precision

(b) Recall

(c) F-score

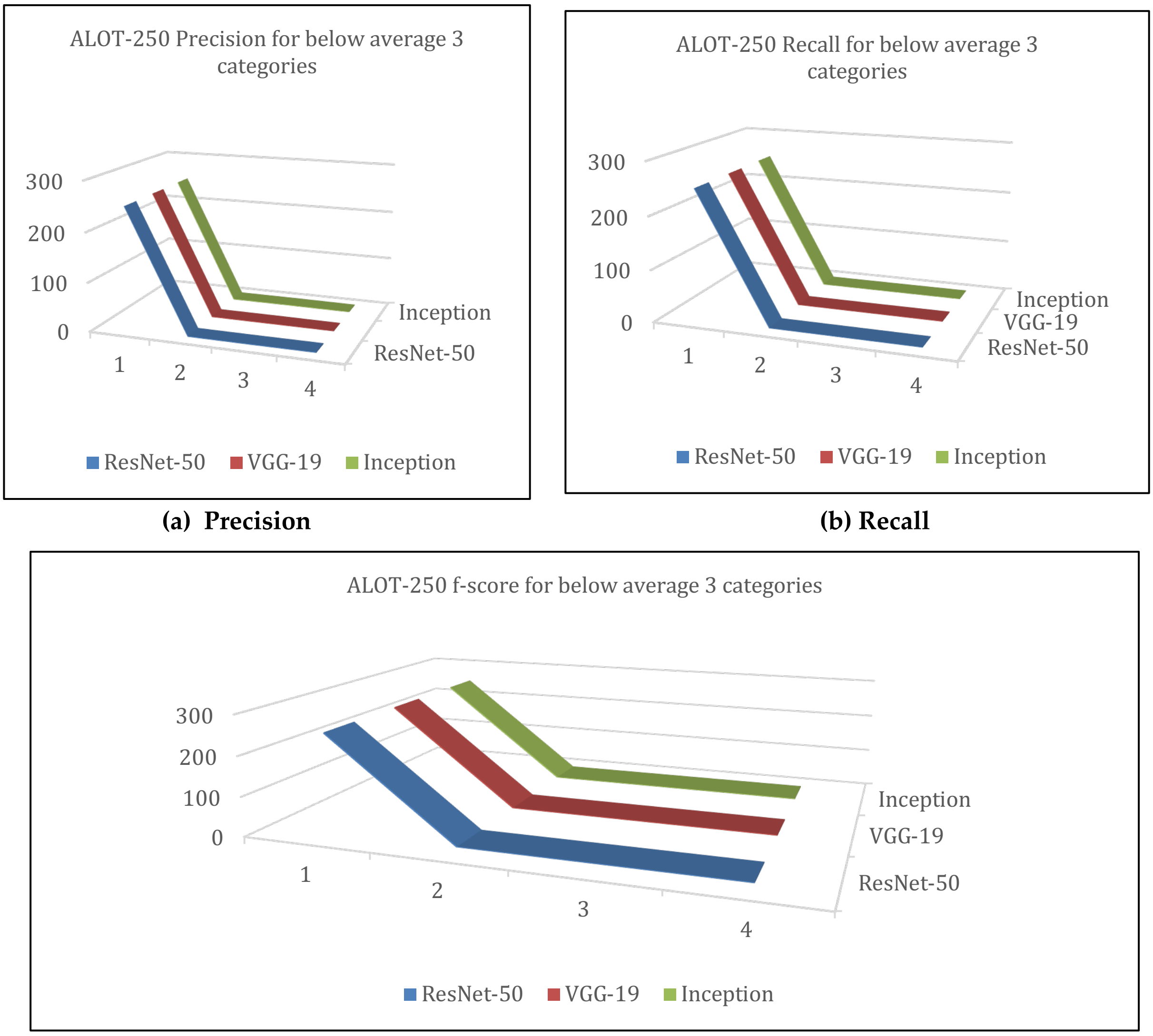
Figure 5. (a) The precision of average 5 categories is outlined for ResNet-50, VGG-19, and Inception. **(b)** The recall of average 5 categories is outlined for VGG-19, Inception and ResNet-50. **(c)** The f-score of average 5 categories is outlined for VGG-19, Inception and ResNet-50.

The graphical depiction in figure 5a showcases the average 5 categories, while table 3 presents the associated numerical data, featuring category names and precision rates. Remarkably, categories such as 126 and 127 achieve a 100% Average Precision (AP) when utilizing the VGG-19, ResNet-50 and Inception architecture. The graphical depiction in figure 5b showcases the average categories, while table 3 presents the associated numerical data, featuring category names and recall rates. Remarkably, highest recall rate is 0.11 and 0.17 while utilizing the VGG-19 and Inception architecture. The graphical depiction in figure 5c showcases f-scores of average 5 categories, while table 3 presents the associated numerical data, featuring category names and f-score rates. The category 126, 127 shows 0.18 f-score using ResNet, VGG-19 and Inception, category 128 shows f-score rate 0.26 using Inception and 125 category shows 0.20 f-score with

VGG-19. Remarkably, highest f-score rate is 0.18, 0.20 and 0.26 while utilizing the VGG-19, ResNet-50 and Inception architecture.

Table 3. F-score, Recall and Precision with ResNet-50, VGG-19, and Inception of ALOT-250 dataset average 5 categories

ALOT-250 Dataset average 5 categories									
	ResNet-50			**VGG-19**				**Inception**	
Category	**Precision**	**Recall**	**F-score**	**Precision Recall**		**F-score**	**Precision**	**Recall**	**F-score**
124	1	0.1	0.18	1	0.1	0.18	0.9	0.11	0.20
125	1	0.1	0.18	0.9	0.11	0.20	1	0.1	0.18
126	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
127	1	0.1	0.18	1	0.1	0.18	1	0.1	0.18
128	1	0.1	0.18	1	0.1	0.18	0.6	0.17	0.26



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

(c) f-measure

Figure 6. (a) The precision of below average 3 categories is outlined for VGG-19, Inception and ResNet-50. **(b)** The recall of below average 3 categories is outlined for VGG-19, Inception and ResNet-50. **(c)** The f-score of below average 3 categories is outlined for VGG-19, Inception and ResNet-50.

The graphical depiction in figure 6a showcases the below average 3 categories, while table 4 presents the associated numerical data, featuring category names and precision rates. Remarkably, categories such as 248 and 250 achieve a 100% Average Precision (AP) along with ResNet-50.The category 249 achieved 70% and 80% precision along with ResNet-50 and VGG-19. The graphical depiction in figure 5b showcases the below average 3 categories, while table 4 presents the associated numerical data, featuring category names and recall rates. The category 249 shows 0.13 recall using ResNet, 250 category shows recall rate 0.25 and 249 category shows 0.14 recall with VGG-19. Remarkably, highest recall rate is 0.25 while utilizing the VGG-19, ResNet-50 and Inception architecture. The graphical depiction in figure 6c showcases below average 3 categories, while table 4 presents the associated numerical data, featuring category names and f-score rates. The category 249 achieved 0.22, 0.23 and 0.29 rates along with ResNet-50,VGG-19, Inception . Remarkably, highest f-score rate is 0.23 and 0.29 while utilizing the VGG-19, ResNet-50 and Inception architecture.

Table 4. F-score, Precision and Recall with ResNet-50, VGG-19, and Inception of ALOT-250 dataset below average 3 categories

ALOT-250 Dataset below average 3 categories									
	ResNet-50			**VGG-19**			**Inception**		
Category	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**	**Precision**	**Recall**	**F-score**
248	1	0.1	0.18	0.8	0.13	0.22	1	0.1	0.18
249	0.8	0.13	0.22	0.7	0.14	0.23	0.2	0.5	0.29
250	1	0.1	0.18	1	0.1	0.18	0.4	0.25	0.31

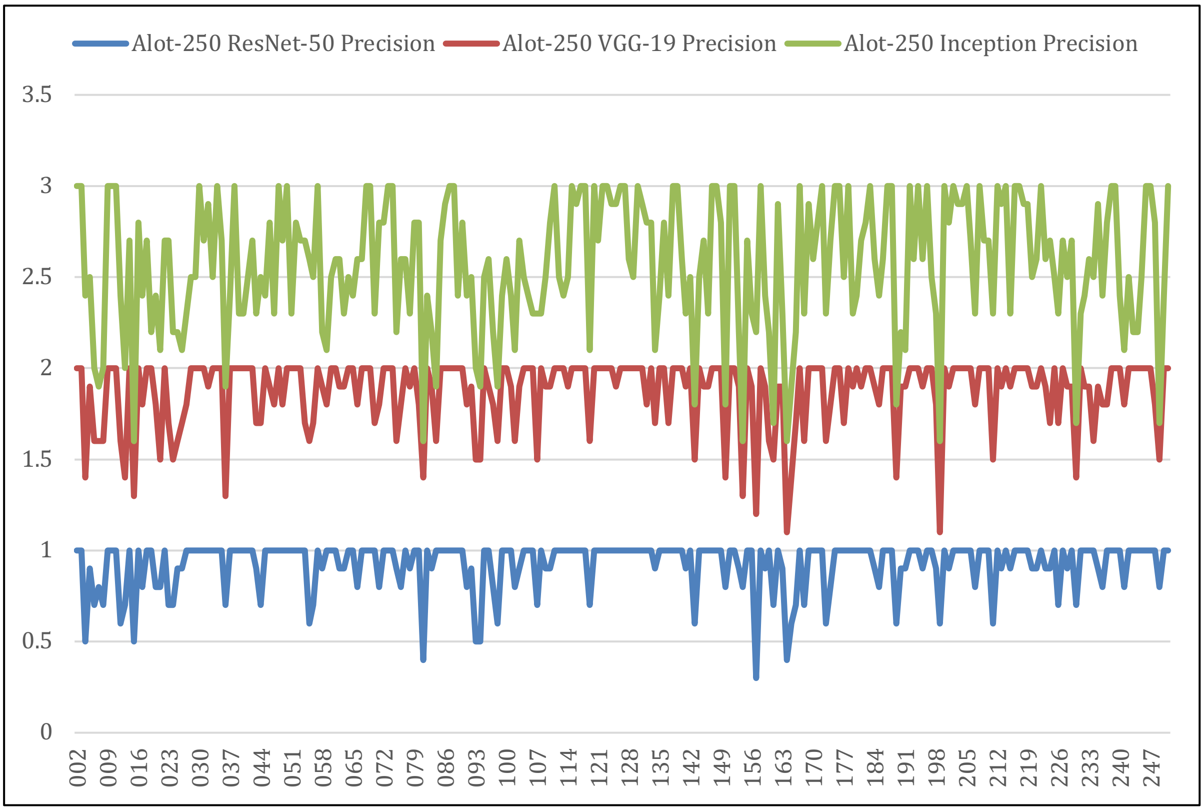


Figure 7. Precision evaluation for ALOT-250

The experimentation results gathered or visualized using a graph to assist comprehension. Figure 7 demonstrates the mean precision of these categories, indicating that VGG-19 and Inception achieve a higher precision rate, while ResNet exhibits a lower precision rate. Highest AP of 1 on some categories (002, 003, 010, 030, 038, 048, 150, 113, 246) along with ResNet-50, VGG-19, Inception architectures.

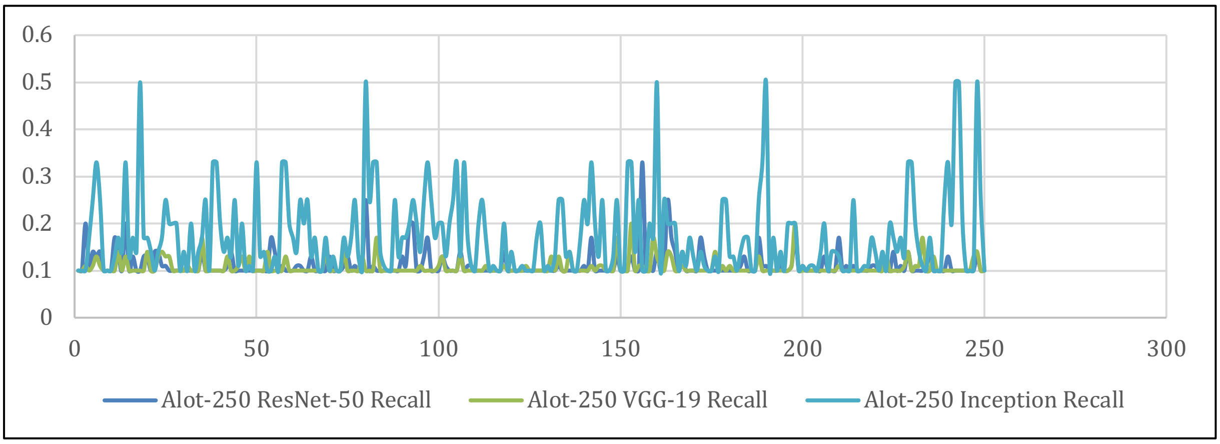


Figure 8. Recall evaluation for ALOT-250

In Figure 8, recall for the classes in the ALOT data set is displayed to illustrate the recall values and their fluctuations, facilitating the evaluation. Inception exhibited impressive performance, with recall values below 30%. Moreover, in the case of ResNet-50, it was noted that the recall for nearly fifty percent of the categories remained around 60%. Highest recall rate 0.33 is observed for categories (083, 084, 098, 143, 241) using inception and for category 157 in ResNet-50.

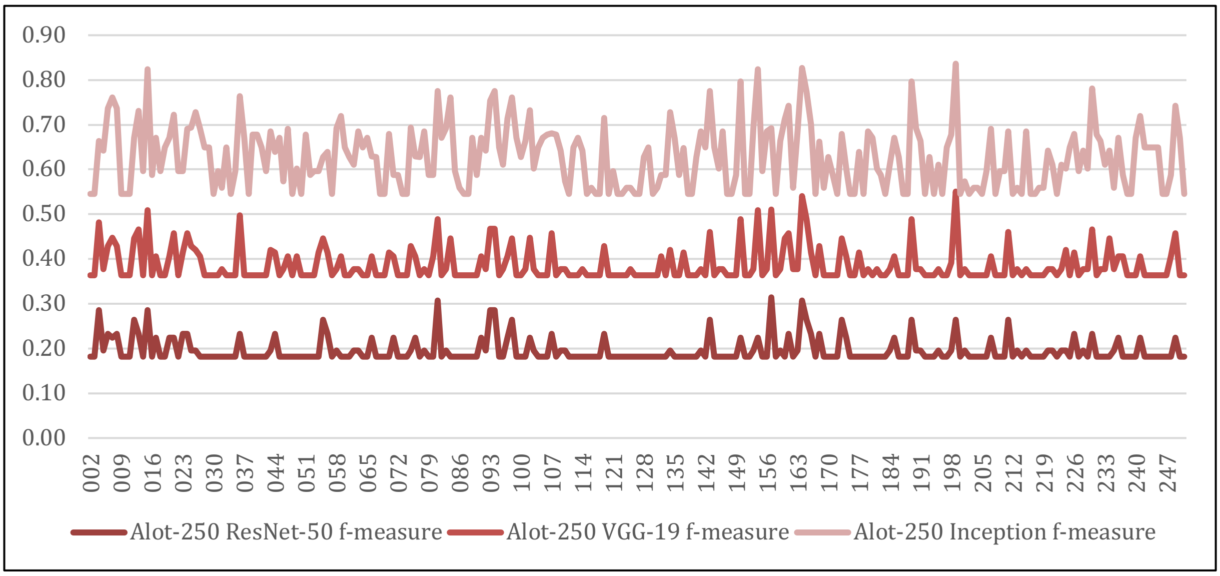


Figure 9. F-measure evaluation for ALOT-250

The F-measure scores for the ALOT-250 dataset using CNN models are illustrated in figure 9. The proposed method demonstrates rates exceeding 70% for several classifications. Highest recall rate 0.31 for several categories 015, 037,135 using inception. Highest recall rate 0.31 observed for category 157 using ResNet-50.

5. Conclusion

Challenges in IT infrastructure and cloud computing have identified quite useful possibilities and new tendencies are also to amplify operational efficiency, security, and sustainability of the processes. One of them is the multi-cloud and hybrid approach which implies the usage of both the public and private clouds to address certain requirements. This Numerous techniques in the Content-Based Image Retrieval field have attracted significant attention and usage for CBIR systems. Nonetheless, the proposed method sets itself apart with its precision and effectiveness compared to others. This model is crucial in detecting, describing, recognizing, and correlating image patterns that accurately depict the true attributes of an image. Novel approach illustrates the image data underwent diverse transformations, including grayscale conversion, pixel intensity analysis, and circular pattern formation, followed by retinal sampling and coefficient formation. Mathematical operations such as differentiation, smoothing, convolution, and max pooling are applied, common in deep learning for feature extraction. A color feature vector was generated and refined using a non-linear function. Steps like mask sizing, kernel induction and exponential adjustment likely pertain to filter application in convolutional neural networks. Overlapped fields are created, with standard deviation applied, possibly to highlight key features. Massive pairing aimed at finding corresponding features between input and reference images, with spatial distance computation to determine distances. Finally, results are aggregated using FV, employing methods like Bag-of-Words (BoW) and K-Nearest Neighbors (KNN) for image classification or object identification. In order to faithfully depict image content, CNN models, such as Inception, VGG-19 and ResNet-50 models, utilize feature integration methods extracted from identifiers or deep learning methods. Testing is implemented with standard benchmarks like 17-Flowers and ALOT-250 datasets. The resulting methods showcase outstanding outcomes throughout diverse image datasets.

References

1. C. Jin and S. W. Ke, "Content-Based Image Retrieval Based on Shape Similarity Calculation," 3D Res., vol. 8, no. 3, 2017, doi: 10.1007/s13319-017-0132-0.
2. "verma2017 local neghbourhood ."
3. R. R. Saritha, V. Paul, and P. G. Kumar, "Content based image retrieval using deep learning process," Cluster Comput., vol. 22, pp. 4187–4200, 2019, doi: 10.1007/s10586-018-1731-0.
4. 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, pp. 1–6, 2019, doi: 10.1109/ICCISci.2019.8716437.
5. K. T. Ahmed, S. Ummesafi, and A. Iqbal, "Content based image retrieval using image features information fusion," Inf. Fusion, vol. 51, no. September 2018, pp. 76–99, 2019, doi: 10.1016/j.inffus.2018.11.004.
6. 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.
7. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies and applications to object detection," Prog. Artif. Intell., vol. 9, no. 2, pp. 85–112, 2020, doi: 10.1007/s13748-019-00203-0.
8. K. T. Ahmed, H. Afzal, M. R. Mufti, A. Mehmood, and G. Y. U. 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," pp. 90351–90379, 2020, doi: 10.1109/ACCESS.2020.2993721.
9. R. Scherer, Computer Vision Methods for Fast Image Classification and Retrieval. 2018.
10. 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.
11. https://ieeexplore.ieee.org/abstract/document/8977568/
12. https://link.springer.com/article/10.1007/s12559-022-10084-6
13. https://link.springer.com/article/10.1007/s10462-023-10478-4
14. M. B. Rao, B. P. Rao, and A. Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features," Int. J. Comput. Appl., vol. 18, no. 6, pp. 40–46, 2011, doi: 10.5120/2285-2961.
15. K. Ivanova, "Content-based image retrieval in digital libraries of art images utilizing colour semantics," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 6966 LNCS, pp. 515–518, 2011, doi: 10.1007/978-3-642-24469-8_62.
16. M. E. Elalami, "A novel image retrieval model based on the most relevant features," Knowledge-Based Syst., vol. 24, no. 1, pp. 23–32, 2011, doi: 10.1016/j.knosys.2010.06.001.
17. Alahi, R. Ortiz, and P. Vandergheynst, "FREAK : Fast Retina Keypoint".
18. X. Wang and Z. Wang, "A novel method for image retrieval based on structure elements' descriptor," J. Vis. Commun. Image Represent., vol. 24, no. 1, pp. 63–74, 2013, doi: 10.1016/j.jvcir.2012.10.003.
19. Irtaza, M. A. Jaffar, E. Aleisa, and T. S. Choi, "Embedding neural networks for semantic association in content based image retrieval," Multimed. Tools Appl., vol. 72, no. 2, pp. 1911–1931, 2014, doi: 10.1007/s11042-013-1489-6.
20. S. M. Butt, "VISUAL FEATURE EXTRACTION FOR CONTENT-BASED IMAGE RETRIEVAL VISUAL FEATURE EXTRACTION FOR CONTENT-BASED IMAGE," no. June, 2014.
21. 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.
22. Z. Zhao, Q. Tian, H. Sun, X. Jin, and J. Guo, "Content Based Image Retrieval Scheme using Color, Texture and Shape Features," Int. J. Signal Process. Image Process. Pattern Recognit., vol. 9, no. 1, pp. 203–212, 2016, doi: 10.14257/ijsip.2016.9.1.19.
23. K. Iqbal, M. Odetayo, A. James, R. Iqbal, N. Kumar, and S. Barma, "An efficient image retrieval scheme for colour enhancement of embedded and distributed surveillance images," Neurocomputing, vol. 174, pp. 413–430, 2016, doi: 10.1016/j.neucom.2015.03.120.
24. K. T. Ahmed and M. A. Iqbal, "Region and texture based effective image extraction," Cluster Comput., vol. 21, no. 1, pp. 493–502, 2017, doi: 10.1007/s10586-017-0915-3.
25. http://growingscience.com/beta/ijds/6908-integrated-multi-layer-perceptron-neural-network-and-novel-feature-extraction-for-handwritten-arabic-recognition.html
26. K. Kanwal, Deep learning using isotroping, laplacing, eigenvalues interpolative binding, and convolved

determinants with normed mapping for large-scale image retrieval, 2021.

27. K. T. Ahmed, Symmetric Image Contents Analysis and Retrieval Using Decimation, Pattern Analysis, Orientation, and Features Fusion, 2021.