scientific reports

Check for updates

OPEN

Efficient CNN architecture with image sensing and algorithmic channeling for dataset harmonization

Khadija Kanwal[1,2]✉, Khawaja Tehseen Ahmad[3], Aiza Shabir[2], Li Jing[1], Helena Garay[4,5,6], Luis Eduardo Prado Gonzalez[4,7,8], Hanen Karamti[9]✉ & Imran Ashraf[10]✉

The process of image formulation uses semantic analysis to extract influential vectors from image components. The proposed approach integrates DenseNet with ResNet-50, VGG-19, and GoogLeNet using an innovative bonding process that establishes algorithmic channeling between these models. The goal targets compact efficient image feature vectors that process data in parallel regardless of input color or grayscale consistency and work across different datasets and semantic categories. Image patching techniques with corner straddling and isolated responses help detect peaks and junctions while addressing anisotropic noise through curvature-based computations and auto-correlation calculations. An integrated channeled algorithm processes the refined features by uniting local-global features with primitive-parameterized features and regioned feature vectors. Using K-nearest neighbor indexing methods analyze and retrieve images from the harmonized signature collection effectively. Extensive experimentation is performed on the state-of-the-art datasets including Caltech-101, Cifar-10, Caltech-256, Cifar-100, Corel-10000, 17-Flowers, COIL-100, FTVL Tropical Fruits, Corel-1000, and Zubud. This contribution finally endorses its standing at the peak of deep and complex image sensing analysis. A state-of-the-art deep image sensing analysis method delivers optimal channeling accuracy together with robust dataset harmonization performance.

Keywords Features fusion, Composite structure, Architectural bonding, Algorithmic channelizing, Deep learning

Digital image processing is an emerging field of computer vision[1]. Computer vision is being applied to examine, understand, and automatically extract valuable information from images. Most of the research being done in this field focuses on vehicle navigation[2], video encoding and robots[3], image retrieval[4], tracking video object recognition[5], video shot retrieval[6], face detection[7], texture classification[8], objects categorization[9], face recognition[3] and image registration[10]. Various stages are involved in image processing and feature extraction at the primary stage is used to extract the local and global features of an image. Local features and global features are the main categories of visual features[11]. Local features of an image refer to region, spatial, and interest points[12] while global features are spatial layouts, colors, shapes, and textures[13]. The major issue with global features is that these are not used for semantic gap reduction. Therefore, global features are unable to represent the complete features of the image. Due to this fact, the global features are not appropriate for partial correspondence of images from the retrieval scheme. On the other hand, local features are very helpful in reducing the semantic gap. To extract features, interest point detectors are applied that characterize the local features of the image to

[1]School of computer science and technology, University of Science and Technology of China, Hefei 230009, China. [2]Institute of Computer Science and Information Technology, The Women University, Multan 60800, Pakistan. [3]Department of Computer Science, Bahauddin Zakariya University, Multan 60800, Pakistan. [4]Universidad Europea del Atlantico, Isabel Torres 21, Santander 39011, Spain. [5]Universidade Internacional do Cuanza, Cuito, Bie, Angola. [6]Universidad de La Romana, La Romana, Dominican Republic. [7]Universidad Internacional Iberoamericana, Campeche 24560, Mexico. [8]Fundacion Universitaria Internacional de Colombia, Bogota, Colombia. [9]Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia. [10]Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea. ✉email: khadijakanwal@wum.edu.pk; hmkaramti@pnu.edu.sa; imranashraf@ynu.ac.kr

overcome the problem of global features. Harris[14], Hessian[15], scale-invariant[16], and affine invariant[17] are mostly used methods for interest point detection. Local and global features of an image are combined to give maximal image contents[18] for object recognition. Content-based image retrieval (CBIR) applies the primitive signatures technique to compute spatial coordinates, edges, shapes, and textures with color channeling for translating the interest points at various interest regions.

Innovative image processing has been emerging as a basic requirement for image retrieval and classification accurately. CBIR scheme is applied for object detection and feature extraction at various levels of application development[19]. Objects, shapes, textures, and color features[20,21] are used to perform CBIR. Several methods have been introduced for image feature extraction based on their contents, textual aspects, and semantic attributes. It is a big challenge in modern research to retrieve relevant images with high precision using deep learning that mostly lacks symmetry for feature description and feature extraction. Convolutional neural networks (CNN) are used for large-scale datasets[22] that make them suitable for computer vision. Image feature extraction and representation play an important role in image analysis[22]. Deep learning is used for feature description and learning[23].

Modern CNN image retrieval and organization strategies are being used due to high performance to cope with many tasks. Furthermore, it is required to fill the gap in image descriptions that are based on human perceptions and features. Semantic gaps are considered human perceptions. CNN works for the extraction of deep features, semantic gap is decreased between low-level features as well as human observation[4]. Fundamentally, deep learning models are used to generate image features, induce convolution and pooling, and introduce new descriptors for features that are analogous at several stages. AlexNet[24], VGG[25], ResNet[26], DenseNet[27] and GooLeNet[28] are reported with competitive results using their structural strength. The model among pool options is the GoogLeNet model which has a special inception module to reduce computational complexity and minimize memory requirements. However, it is not obvious to achieve improved performance using CNN all the time. To overcome this issue, researchers introduced an efficient feature fusion method with a combination of features at the lower level by GoogLeNet fully connected features (GLFCF) and dot-diffuse block truncation coding (DDBTC) at the higher level. Furthermore, a vector of locally aggregated descriptors (VLAD) is employed for the excessive features reduction dimension[29]. The resNet model uses an additional hundred convolution layers. ResNet requires an enormous volume of computational time for training and classification steps. It is a big challenge to classify these systems in real-time applications[26]. To deal with this issue, the graphics processing units (GPUs) are used to improve processing time for classification and training[30].

The ResNet architecture presented an idea for connecting early layers to the later layers, which is called identity mapping. Currently, DenseNet pushed this technique to a great extent by creating connections among all sets of layers in the dense block. DenseNet architecture can be classified as a general form of the ResNet architecture because DenseNet creates more identity mappings between layers. ResNet model identity mappings add identity mapping features from the early layer to the recent layer for information preservation and gradient, while the DenseNet model combines the knowledge from all ancestor layers to concatenate them using a new feature. This is achieved by increasing the filter size of the convolutional layers when the layer goes deeper[27]. CNN is the most widely used deep-based learning model to encourage natural graphic perceptions for living beings. CNN architectures are utilized at higher-level abstractions with further deep learning architectures to syndicate several non-linear transformations. Better performance has been achieved by CNN to solve various problems for domains such as natural language processing[31,32], speech recognition[33], and object recognition[34]. The CNN architectures are offered to improve the workflow of image retrieval[31,34,35]. Moreover, deep learning-based feature extraction procedures provide useful offerings in multimedia content processing. CNN features efficiently attained the performing tasks in the field of computer vision and image retrieval methods[36]. This is done by a comparison of local features with CNN features that contain scale-invariant feature transform (SIFT) constant manners under different image revolutions. So, it is considered a phenomenon visible at deep learning features combining level with descriptors for primitive level. CNN model is considered a tensor that inputs an array of multi-dimensions, and the result is collected at a high-level aspect object structure. The probability function is applied as input data for proper regression and classification. It is highlighted how CNN can optimize feature extraction by simultaneously categorizing it into a single task such as a learning block in the training phase for CNN eliminating the requirement of feature extraction[37].

Modern CNN classification methods experience substantial difficulties when attempting to generate unified and optimized feature vectors that function across diverse datasets exhibiting different semantic contexts and structures. The processing capability of various existing methods remains limited for both color and grayscale data which leads to inadequate performance when handling datasets with diverse characteristic features. The majority of existing solutions work with single CNN architectures while lacking integration methods to harness the combined assets of multiple models. The absence of cohesive integration between frameworks reduces the potential to extract complete image features from input data. CNNs show restricted effectiveness when processing anisotropic noise while struggling to detect important fine details including peaks, ridges, and junctions needed for precise image analysis. The existing processing methods lack effective mechanisms for detecting and processing these image features reliably. Pattern recognition using CNN models shows weakness in adaptable learning across semantic challenges within different structured datasets causing performance deterioration in new evaluation sets. Our research established a new framework by structuring DenseNet with ResNet-50, VGG-19, and GoogLeNet through architectural binding and algorithmic routing technology. Features maintain consistency across different datasets through this approach together with corner straddling methods that resolve anisotropic noise and extract fine-grained features via curvature-based calculations. The proposed framework offers exceptional performance while maintaining generalization ability across diverse benchmark datasets by solving existing gaps.

This novel research is an application of the local features that are collected and strengthened by DenseNet, VGG-19, ResNet-50, and GoogLeNet models with presented algorithms for image feature extraction. The presented model introduces a novel approach that incorporates straddling, autocorrelation, factoring, thresholding, junctioning, regioning corner response, signature influencing, dataset harmonizing, image indexing, and deep sense image classification combined to create detail for deep improved image contents. Deep learned features are the footprints of effectively combined signature influencing connected with architectural algorithmic elaborations and the algorithmic channeling with appropriate coefficient contributions and parallel joints of color displacements. This proposed model aims to obtain compact and efficient image feature vectors to deeply learn the images parallel at colored and grey integrations along with CNN signatures, with channeled algorithm integrations to have uniformity at diverse datasets and semantic groups. In particular, we focus on the following objectives.

- **Proposed Approach**: We expanded on the choice to integrate multiple CNN architectures DenseNet, ResNet-50, VGG-19, and GoogLeNet to leverage their complementary strengths in feature extraction, which enhances the model's versatility and performance on complex image datasets.
- **Multi-Channel Processing**: The use of both color and grayscale channels allows the model to capture both texture and structure, improving accuracy in datasets like Caltech-101 and 17-Flowers, where color and texture are important for classification.
- **Addressing Feature Fusion Gaps**: We introduced the fusion of local-global vectors and region features to improve the model's ability to handle diverse datasets, offering a more balanced and robust feature representation compared to traditional single-feature extraction methods.
- **Noise Handling**: We strengthened the rationale behind handling anisotropic noise by using auto-correlation to generate curvature-peaked, ridge, and peaked responses, which improves robustness to noise, particularly in datasets like COIL-100 and Corel-1000.
- **Channeled Algorithm Integration**: The channeled algorithm fusion enhances feature representation and retrieval accuracy by enabling more effective communication between different architectures, improving performance across diverse datasets.We obtained results by encapsulating the image patching with straddling at the corner and isolated responses, which resulted in anisotropic noises with the observation autocorrelation for shape parameters to be output in curvature peaked, ridge, and peaked responses to be further computed for factoring, thresholding, junctioning, and regioning. The parallel response is also obtained from the merged proposed channeled algorithm with DenseNet, ResNet-50, VGG-19, and GoogLeNet incorporated with spatial heads generated from displaced color channels to form architectural bonding. The image channeling is achieved on Color channels that are applied to combine the color coefficients and L1 and L2 normalization is used at these RGB color channels, and incorporated with formed signatures to show the compacted massive feature vectors. These robust deep features are induced to detect complex, cluttered, textural, spatial, mimicked, overlay, tiny and large images. These feature vectors are treated as input to BoW architecture for efficient image classification and image retrieval. This proposed philosophy is experimented on standard benchmarks including Caltech-101, Cifar-10, Corel-10000, Cifar-100, Caltech-256, 17-Flowers, COIL-100, FTVL Tropical Fruits, Corel-1000, and Zubud, and reported significant results.

The contents of this article are arranged as follows. "Related work" Section provides concise existing research work about CBIR and convolution neural networks (CNNs) with deep learning. The presented methodology with algorithms is explained in "Methodology" section. "Experimentation" section elaborates on the experimentation, also results with tables and graphs, and comparisons with existing designed research approaches. In "Conclusion" section, the main findings and contributions of the proposed methods are summarized with a discussion of future work.

Related work

Modern research contributions based on CNN focus on dense prediction problems namely image generation, semantic segmentation, and style transfer. The residual network significantly increased image classification accuracy and has been extensively used. CNN models have improved progress and provided solutions to many problems related to the field of computer vision. Harris and Laplacian based surrounded combined support vector machine (SVM) based feedback method is introduced with several techniques and algorithms proposed to match the image contents with accurateness[38]. The corner detection descriptors are used for corner extraction from images wherever the density ratio is applied to obtain prominent regions for image distinguishing parts. The shape retrieval from images with color channeling information is combined to identify different important regions. The Harris corner detector is also used with the bi-directional decomposition technique for CBIR[39] and the Harris corner detector is applied to detect corners, and the bi-dimensional empirical mode decomposition (BEMD) method is used to extract the edges[40]. For this, extracted features are joined to achieve image retrieval by using datasets with these two methods. An image recognition-based method is proposed with a speeded-up robust features (SURF) descriptor. The SURF is an interest point detector and descriptor of images as presented in[41,42]. SURF detector is applied to extract requisite images and matching feature points from the image. Multiple instances learning SVM and feature extraction using a SURF detector are presented for the image retrieval and classification in[42]. Using Fisher vectors (FV) a technique is proposed in[43] to extend bag-of-visual words (BOVW) that is based on intermediate image depiction applied for image classification. Another method proposed in[44], is an image retrieval system based on contents and transfer learning from a CNN trained on wide image datasets.

The shape, color, spatial coordinates, edges, and texture are considered low-level attributes of the image. Feature merging is a very useful approach that is used in CBIR to improve performance. The fusion of color and texture features have been presented in[45] to extract the local features as feature vectors. Moreover, in[46],

a technique is proposed that is based on three main steps: feature extraction, performance evaluation, and equality for similarities. For color feature extraction, color moments (CM) are used; and discrete wavelet and Gabor wavelet transform techniques are applied to extract texture features. Furthermore, a descriptor is used to increase the effect of feature vector presentation also identified as color and edge directivity descriptor (CEDD). A useful innovation is reported in[47], which provides a transformative recovery technique (CBIR-GAF) based on the combination of four global descriptors. Special image retrieval problems are reported in[45] for specifically designed domains using CBIR. To overcome image retrieval and classification problems, a method is introduced based on various techniques such as retention for CNN. In this approach, image feature extraction is performed with fully linked layers after retaining of proposed CNN model functioned as feature vectors for each image. To compute the distance between the query image and feature vector to nearest neighbors most same image from the repository, Euclidean distance (ED) is applied[48]. Many researchers reported semantic gap problems between low-level image features and human attention. To overcome the semantic gap, another useful technique is introduced in[49] with the combination of CNN and sparse representation. Experiments are conducted on MPEG7, Corel, and ALOT datasets for image retrieval.

CNN is focused on removing image credits at the final layer by applying solitary CNN design commonly through less quantization methods which bounds the utilization of central convolutional layers to recognize the image neighborhood measures. An architecture is applied with deep CNNs that are effectively pre-trained on a great basic image database to improve the performance of CBIR. To reduce the image features, a capable bi-linear root pooling technique is suggested which is useful for low-level dimension pooling layers to dense due to excessive discriminative descriptors of the image. A novel image interpretation technique is proposed with machine learning and deep learning algorithms. This approach has provided better precision with a combination of AlexNet CNN, local binary pattern (LBP) descriptors, and histogram of oriented gradients (HOG). Transfer learning is explained in[50] which is used in several fields of science; remote sensing convolution is capable of applying to the natural image that is organized by CNN. A spatial division network is introduced to recognize bounding boxes only with delicate observation. A principal component analysis (PCA) technique is used for image classification with deep belief networks (DBN) in[51]. A supervised hashing-based novel approach is presented for image retrieval that filters the banks using block histogram and performs pooling with binary hashing and indexing of image representation with CNN[52].

CNNs have been explored by several advanced methods to attain improved performance for different applications. CNN approaches are proficient tools to attain improved performance in image indexing and retrieval. In recent times, tendencies have focused on comprehending deep neural networks (DNN) improved performance at numerous convolutional levels. CNN multiple layers extract different features with the local approachable fields by replicating cells in sub-sampling and visual cortex layers. Furthermore, the layer for sub-sampling filters response to attain scarcity invariances in complex cells functions for the visual cortex. A different technique is proposed to determine the slight changes in results of image indexing[25]. Moreover, networks could be extended to add novel layers and apply novel algorithms for computer vision[53]. In[54], a features extraction technique is presented for image retrieval and classification using CNN. A deep learning-based effective approach is presented to extract image features using ResNet-50. This method is applied on seven stages namely sampling to find symmetry, FAST score-based suppression, different scaling levels, space-based sampling, feature reduction, smoothing, and filtering in[55].

In[56], an approach is presented with the fusion of GoogLeNet, ResNet, and VGG models using an interconnection. SVM and random forest-presented algorithms are used for image retrieval and classification. Experiments are conducted to improve performance for the Standford 40 actions. The presented VGG-19 deep neural network-based DR model is the outstanding AlexNet architecture and introduces SIFT for computational time and sorting precision. For this, the stemming method for image labeling[57], AlexNet, and GoogLeNet are used to detect the objects and recognition on the ImageNet dataset using DNN. Moreover, GoogLeNet, ResNet, VGG and AlexNet models are useful to recognize scenes of places365 dataset. The HOG method is applied to detect the number of persons. A model is proposed for the distinctions and its performance is compared with the newest CNN variants including GoogLeNet, ResNet, VGGNet and AlexNet on the BelgiumTS dataset in[34]. Another approach is introduced to enhance the processing speed and results recognized objects between CNN[58]. Feature selection is achieved through principal component analysis and SVD by FC layers that confirm the accuracy level for image classification. This method is combined with GoogLeNet, ResNet-50, and VGG-19 with a maximum response based on convolutional Laplacian scale object and Eigenvalues with color channels to retrieve the images efficiently[59]. This technique supports the scaled object features that are normalized with BoW of fast image retrieval and classification.

The study[60] investigates perception-guided U-shaped transformer networks built to provide 360-degree no-reference image quality assessments. The detection method demonstrates that perception-based learning methods provide essential value because they support the ongoing effort to advance image quality analysis technology. Similarly,[61] demonstrates a new transformer-based system that detects and tracks image defects within infrastructure maintenance settings. The approach of using transformer architecture to detect defects shares essential principles with the proposed image evaluation method through the integration of deep learning networks and channeled algorithms to extract and refine image features.

The research[62] explores techniques for boosting image quality in dim lighting as a solution to typical imaging sensor issues. The goal is to enhance images through algorithmic manipulations for feature refinement requires framework robustness across various imaging conditions and aligns with this study's objectives. The study[63] presents an innovative iterative process to combine self-organizing matrix entanglements for remote sensing image classification. This work utilizes the combination of DenseNet and ResNet-50 as neural network architectures but concentrates on implementing these advanced classification techniques to analyze image data.

Along the same lines,[64] focuses on implementing state-of-the-art classifier methods for analyzing image data. The study investigates resource-efficient image-sensing techniques by using binary optical computing methods. Deep learning demand requires resource-efficient methodologies, which makes this insight more appropriate now than ever. Our system uses minimal-sized efficient image feature vectors to follow current computing standards, which demand reduced operations while delivering superior accuracy outcomes.

Two other studies including[65] and[66] reveal innovative image geometry adjustment techniques vital for improving image quality before analysis while echoing our work on elaborating consistent image data collections.

In the context of image fusion, FreqGAN presented in[67], demonstrates a methodology to merge infrared and visible images through frequency domain-enabled generative adversarial networks (GANs). The research matches this direction, which uses diverse image data sources while applying network-based fusion practices to boost feature extraction capabilities. Likewise,[68] demonstrates the potential of generative models to obtain significant features from complex image datasets while replicating our deep learning approach that emphasizes sophisticated feature extraction methods. In addition, studies outside the direct scope of image sensing, such as[69] and[70] expand the networked systems and communications field while highlighting the critical importance of machine learning adoption in modern communication technologies. The research shows that adaptive algorithms alongside diverse computational methods are essential components, which inspired the current framework development through algorithmic channeling.

Two academic publications,[71] and[72] reveal how real-time systems benefit from context-aware adaptive algorithms. The framework prioritizes semantic groups while demonstrating operational effectiveness throughout multiple imaging conditions, which range from mobile to real-time formats. Two studies explore system optimization within crowded spectrally challenged V2I communication networks as well as UAV-to-ground UWB channels affected by built-up areas and airframe implementations. The requirement to optimize image sensing algorithms matches the need to achieve robust high-performance detection across multiple real-world operational scenarios.

The presented approach is designed with the most suitable technique to classify and retrieve the images. This method presents a novel approach that incorporates straddling, autocorrelation, factoring, thresholding, junctioning, regioning corner response, signature influencing, dataset harmonizing, image indexing, and deep sense image classification combined to create detail for deep improved image contents. We obtained results by encapsulating the image patching with straddling at the corner and isolated responses, which resulted in anisotropic noises with the observation autocorrelation for shape parameters to be output in curvature peaked, ridge, and peaked responses to be further computed for factoring, thresholding, junctioning, and regioning. The image channeling is achieved on color channels that are applied to combine the color coefficients and L1 and L2 normalization is used at these RGB color channels, and incorporated with formed signatures to show the compacted massive feature vectors. This proposed philosophy is experimented on standard benchmarks namely Caltech-101, Cifar-10, Caltech-256, Corel-10000, 17-Flowers, Cifar-100, COIL-100, FTVL Tropical Fruits, Corel-1000 and Zubud, and provided astonishing results.

Methodology

This section provides the details of the proposed approach and its working mechanism. All sub-modules are described in detail. This study focused on the following aspects:

- **Novel Integration Techniques and Modifications:** The study uses novel image formulation and synthesis functions. Specifically, we focus on the integration of DenseNet, ResNet-50, VGG-19, and GoogLeNet using channeled algorithm integration with displaced color channels for enhanced image representation and retrieval performance.
- **Supporting Results**: To substantiate these optimizations, we have included a series of quantitative and qualitative results that demonstrate the improvement in performance.A detailed explanation of the channeled algorithm integration, color displacement, and signature harmonization processes is provided in the subsequent sections. We also emphasize how these modifications contribute to semantic uniformity and improved image retrieval.

Grey shaded images

In the first step, the colored query image is converted into a white and black image using the formula presented in Eq. (1)

$$Image = RGB2grey\ level \tag{1}$$

Equation (1) is used for the conversion of color images to gray-level images. The gray-scale image is used to reduce complex images such as shades, texture, contrast, edges, shadow, shapes, etc. with considering colors. Normally, the greyscale image is more important than its colored image due to one color processing. Another purpose of applying the conversion technique is that colors are arranged by humans, but query images are arranged with a more complex system. After applying conversion on colored images into greyscale images, feature vectors are created by using descriptor algorithms and process of interest points detector as illustrated in Fig. 1.

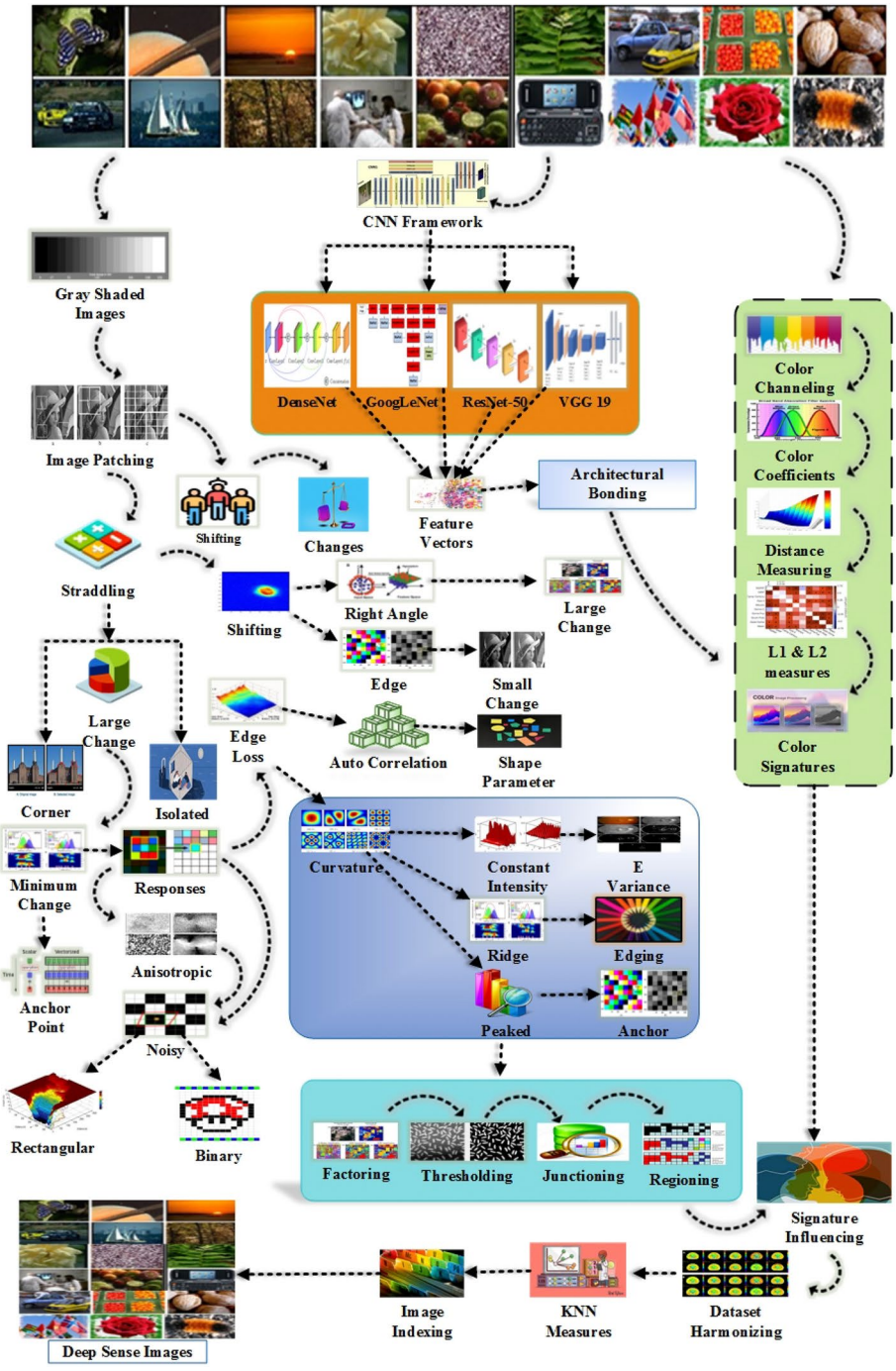


Fig. 1. Architecture of proposed model.

Algorithmic channelizing

At this step, corner detector functions of Moravec's are considered as a local window in an image and determine the normal changes of the image intensity which is the result of shifting the local window to a minor extent in different directions. The most important three conditions considered are

- If the window patch is almost constant in its intensity, then total shifts will outcome in simply a small change.
- If a window straddles an edge, then a shift alongside the edge will outcome in simply a minor change however a shift vertical to an edge will outcome in a big change.
- If the window image patch is an isolated or corner point, then total shifts will outcome in a big change. A corner could be detected using searching when the lowest change formed with shifts is big.Considering the above-mentioned three conditions, Eq. (2) is defined. Suppose W is denoted as the image intensities, the change C is formed with a shift (a, b) is defined as in[8]

$$C_{a,b} = \sum_{i,j} l_{s,t} |W_{a+s,b+t} - W_{s,t}|^2 \tag{2}$$

Here, l is specified for the image window. This is the unity in a definite rectangle region. The shifts (a, b) are supposed to include $\{(1,0),(1,1),(0,1),(-1,1)\}$. Therefore, the corner detector of Moravec's can easily look for the local maxima in minimum$\{C\}$ above a few thresholding values. Moravec's corner detector performance is measured to test the image and performed comparison in various predefined algorithms. The Moravec's operator has suffered from many issues; these issues are defined below with their proper corrective measures. Algorithmic channelizing is also described in Algorithm 1.

1: Step-1: 3D curve representation
2: Step-2: t=Tracked edging
3: Step-3: w = Wire-framing
4: Step-4: w$\leftarrow d_i$ = deli-mated image
5: Step-5: edging <— Canny filtering
6: Step-6: $H_s \leftarrow Hysteresis \leftarrow S\,purs$
7: $H_s \leftarrow C_{th}$
Where C_{th} is close threshold
8: Step-7: Result $\leftarrow \mathscr{L}$
Where $\mathscr{L}$: Edge topology enforcement

Algorithm 1. Algorithmic channelizing.

Anisotropic responding
The responses are anisotropic due to a discrete set for shifts at each forty-five degrees measured, small shifts could be protected to perform an analytical expansion for shift origin[8]

$$C_{a,b} = \sum_{h,g} l_{s,t}[W_{a+s,b+t} - W_{s,t}] = \sum_{h,g} l_{s,t}[aA + bB + L(a^2, b^2)]^2 \tag{3}$$

In Eq. (3), the first gradients are approximated in[8]:

$$A = W \oplus (-1,0,1) = \frac{\theta W}{\theta a} \tag{4}$$

$$B = W \oplus (-1,0,1)^R = \frac{\theta W}{\theta b} \tag{5}$$

Hereafter, C can be written as for the small shifts in[8]

$$C(a,b) = Xa^2 + 2Zab + Yb^2 \tag{6}$$

where

$$X = A^2 \oplus l$$
$$Y = B^2 \oplus l$$
$$Z = (AB) \oplus l$$

Noisy response generation
The response is noisy due to the window beginning rectangular and binary - usage of smooth rounded window, suppose a Gaussian[8]

$$l_{s,t} = exp - (s^2 + t^2)/2\phi^2 \tag{7}$$

Edges losing
The response of operators also ready for the edges due to a minimum of C takes into account again formulation of the corner measures to use the dissimilarity of C with direction for shift in[8].

$$C(a,b) = (a,b)N(a,b)^R \tag{8}$$

Here 2×2 matrix N is defined as in[8]

$$N = \begin{bmatrix} X & Z \\ Z & Y \end{bmatrix} \tag{9}$$

In Eq. (8), C is closely correlated to the function of local auto-correlation, using N relating its shape at the origin.

Suppose w and k are the eigenvalues of N. Here, w and k are proportionate to the main curvatures for the function of local auto-correlation and use of a rotational invariant definition of N[8]. Three conditions are required to consider

a) The local auto-correlation function is flat when both curvatures are small so that the image region is almost constant in intensity.
b) The local autocorrelation function is ridge shaped when one curvature is highest and the other is lowest so that the shifts along with the edge due to minor change in C indicated as an edge,
c) Local autocorrelation function is sharply peaked when both curvatures are highest so that shifts of any direction are increased C indicated as a corner.

Signature influencing

Edge classification and corner regions are required, as well as, response or measure the quality of edge and corner. The response size is applied to choose the isolated corner pixels and thin the edges. Suppose H is used to measure of corner response that is required as a function of w and k only for the rotational invariance. Determinant (N) and Trace are used in the formula to avoid the explicit eigenvalues decomposition for N in[8], therefore

$$Trace(N) = w + k = X + Y \tag{10}$$

$$Determinant(N) = wk = XY - Z^2 \tag{11}$$

Suppose a formula for the corner response in[8]:

$$H = Determinant - nTr^2 \tag{12}$$

H is used for positive in corner region pixels, minor in hit flat region, and negative in edge region pixels. To improve the contrast in all possible cases is necessary to increase the magnitude of response. The flat region is represented by Trace decreasing a few selected thresholds. A corner region is chosen such as a nomination of a corner pixel when the response is an eight-way local maxima, corners are identified in the testing image. In the same way, edge regions are considered to be edges when responses are local minimal, and negative in either a or b directions, whether the first gradient magnitude in the a or b directions correspondingly is the higher. This results in thinned edges. The classification of raw corner or edge classification with black representing the corner regions and gray, thin edges.

By using high thresholds, low thresholds, and edge hysteresis is carried out and is able to enhance the edge continuity. These types of classifications according to the result in a five-level image including 2 edge classes, 2 corner classes, and background. Next processing (same as junction completion) removes short isolated edges and the edge spurs, and bridges small breaks in the edges. The resultant is the continuity of the thin edges that usually terminate in corner regions. The edges terminator is connected to corner pixels that exist in corner regions to formulate a linked edge-vertex graph. This is very important when several corners in the bush are not linked to edges such as they exist in essence texture regions. Many edges and corners are directly matchable. In Algorithm 2, signature influencing is also described in detail.

1: Step-1: Let $Gs1 = Pi + Si$
   //Where $Gs1$ is staged-1 grey level image,
   //Pi is patched return,
   //Si is straddled version of i
2: Step-2: $Gs1 \leftarrow Sci \rightarrow$ (Change affected)
   //Sci is shifted changed caused affected,
   //Double arrow shows followed impact.
3: Step-3: $B + R \leftarrow NARI + APMC$
   // Where B = Binary, R= Rectangle,
   // $NARI \rightarrow N$ = Noisy, A =Anisotropic,
   // R = Response, I = Isolated
4: Step-4: Apply thresholding followed by factoring
5: Step-5: By using step-3 and step-4, apply regioning with junctioning

Algorithm 2. Signature influencing.

KNN formulation impacts

In the next step, the BoW model is applied to retrieve and index the images. The BoW or Bag-of-visual-words (BoVW) architecture is employed by SVM for image classification and retrieval. This model applies occurrences as features rather than the binary matching of class by class. Furthermore, BoW uses the KNN method that saves the recent instances classified based on the similarity and efficiently generates the outcomes. BoW strengthens with local patches to illustrate the image that deals with numeric vectors with the presented model. Moreover, these numeric vectors are used as candidate descriptors to handle and scale the differences with invariances which are difficult to achieve in classification systems at the binary level. BoW is an efficient and effective method due to its clustering modeling and code-words modeling wherever learned patches are used to map code-words with clustering. The BoW is a dominant way out against other existing models. The BoW demonstration is a histogram depiction with every local descriptor that is assigned to the visual word. For offline training, perform $\{d_1, d_2, ..., d_u\}$ of u clusters trained by the k-means. Histograms for the local descriptors are created to construct a representation of fixed length using u bins of an image and determined by the mapping of all descriptors assigned to the nearest cluster. The frequency of the inverse document is used with the inverted list to compare efficient BoW representations. The BoW images are classified and indexed, and relevant image outcomes are explored from the visual BoW databank and then show the retrieved results.

Color channeling

At this step, RGB images are supposed like color components that are designed as a color channel to denote those features. These color channels are the carriers for the basic colors to present the image features. The importance of the algorithm is that it correspondingly collects the color coefficients along with the gray level strengths. The model performs spatial mapping for these color coefficients to reveal the contents of deep images. Coupling of color information using gray level values produces a high representation of image contents. The color information requires classic objects and also their placement by spatial coordinates resolves the similarities which are intensive in the presented model to obtain better average precision and average recall results. The basic rule of an algorithm is to form the segments for the image based on information, combined from scenes and objects for the image. Generally, the segmentation process is based on the object, label, and scene process of every pixel resultant for these objects specifically in the colored image are measured with difficulty. To overcome this problem, the L2 norm is selected in the presented model. In Eq. (13), the crossing point reflection in the blurred object is defined with the function[7].

$$P(\phi,\gamma) = \begin{cases} G_o(\phi)F_o(\gamma)Metal & (i) \\ G_o(\phi)f_o(\gamma) + G_g(\phi)F_q(\gamma)K.F & (ii) \end{cases} \tag{13}$$

where P is the demonstration for photometric angles and to measure the wavelength, ϕ is applied. γ is denoted by the reflection surface and g demonstrates the body reflection. $K.F$ is homogeneously used for dialectic objects. The body reflection of $K.F$ and reflected surface O of the homogeneous object is linked only in one direction. $G_O(\phi)$ is used to define the sensor values brightened for the objects which are dialectic. The value for reflection of brightened objects is larger than zero and is zero for the blurred objects. To compute spectral power effects on the illumined surface in the image. The following equations are used to calculate sensor values at every (q, d)[9,73]

$$J_r(e,c) = \int_{\phi} Đ_r(\gamma)N(\gamma)H(\phi)P(\gamma)d\gamma \tag{14}$$

$$= H(\phi)\int_{\phi} Đ_r(\gamma)N(\gamma)P(\gamma)d\gamma \tag{15}$$

In Eq.(14) and Eq. (15), ϕ is used to reflect the scene geometry in location (e, c). Generally, the L1 norm is measured by coordinates (e, c) for the color space.

Consider a set $U = (J_0, ..., J_{m-1})$ is represented by the direction line so $L2$ norm is described in Eq. (16)[9,73]

$$\check{U} = \frac{J}{\sqrt{J_0^2 + ... + J_{m-1}^2}} = (\check{U}_0, ..., \check{U}_{m-1}) \tag{16}$$

Suppose

$$Ð_r = \int_{\phi} Ð_r(\gamma) N(\gamma) P(\gamma) d\gamma \tag{17}$$

Then from Eq. (15)

$$\check{U}(q, d) = \frac{H(\phi) E^i}{\sqrt{H(\phi)^2 (E_0^2 + ... + E_{m-1}^2}} = \frac{E^i}{\sqrt{(E_0^2 + ... + E_{m-1}^2)}} \tag{18}$$

$\check{U}_r(q, d)$ is based on input brightness, sensor and reflective result; it is autonomous to ϕ. L1 norm is employed to generate color coordinates that are recognized as the chrome coordinates[11,74], accordingly,

$$\check{U} = \frac{J}{J_o^2 + ... + J_{m-1}^2} \tag{19}$$

L2 normalization coordinates are the same as the L1 norm which does not completely disclose the scene geometry. The presented model uses the L2 norm as an alternative to the L1 norm because to normalization of L2 is an intensive optimization for mean cost rather than the median cost. Generally, the error result is low in the L2 norm with constrained outliers. The main problem of the L1 norm is restricted differentiations because of protecting outfitting and sparsity execution. L2 norm has overcome this problem and demonstrates invariance by enhanced coverage. L2 norm has also one more benefit such as its natural surrounds of squaring input which is a closed form whereas L1 normalization pair-wise is the absolute function; accordingly, L1 normalization is a more expensive computation than the L2 normalization. In L1 normalization, two various resources the distance among the coordinates are based on online location. Two textural dependent materials by $\varepsilon 1$ angle where it is represented angle among the resources. For the Euclidean distance h_1 and L1 normalization equation is defined[73]

$$h_1 = \sqrt{2} \left(\frac{s(\varepsilon + \varepsilon_1)}{(1 + s(\varepsilon + \varepsilon_1)} - \frac{s(\varepsilon_1}{1 + s(\varepsilon_1)} \right) \tag{20}$$

For L2 norm, the Euclidean distance h_1 as in following Eq. (21)[73]

$$h' = n \times J \left(\frac{\varepsilon}{n} \right) \tag{21}$$

Here, h_1 is based on ε, $n = 2$, and J is employed to calculate sine. For the maximum dimensional sensor space, the statement is similar. Then, it is noticeable that the two colors on the surface are the same even though the first point is highlighted directly and the other point is shaded[12]. The two contrary properties by L1 norm connected to calculate normalized colors have been considered. The proportions for color in the sensor space have a zero point $J_0 = J_1 = ... = J_{m-1} = 0$. Algorithm 3 shows the steps for dataset harmonizing.

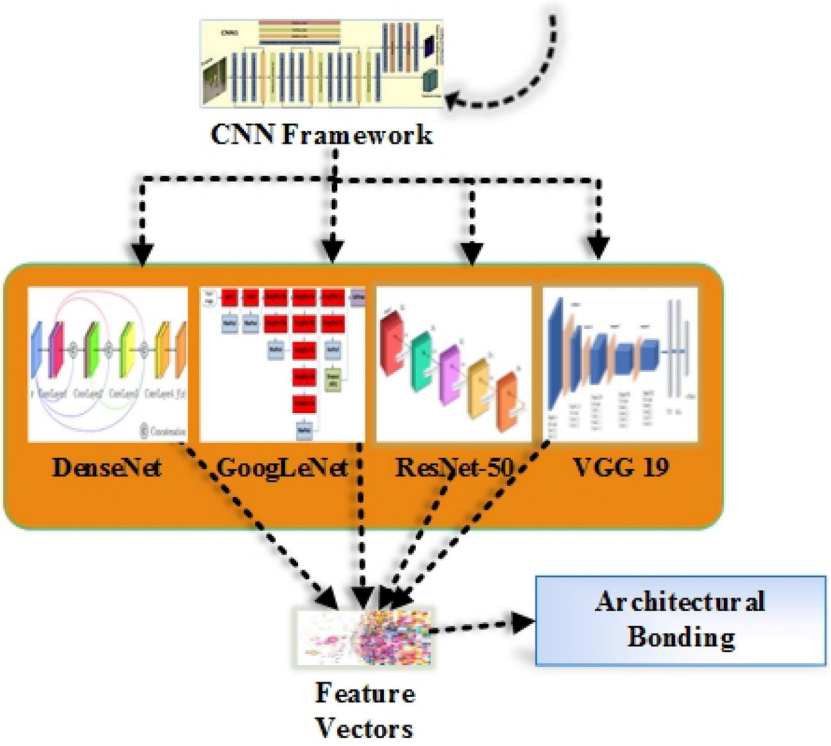


Fig. 2. Architectural bonding.

1: Step-1: $DS = IR + RI + IC + IL$
   //Where S = Size, R= Resolution,
   //C=Color-values, L = Layers
2: Step-2: $A = DS \leftarrow Sg \leftarrow NCS$
   //Where DS is impacted by Sg: Semantic groups
   //And NCS : Number of classes, A is accuracy
3: Step-3: $NCS \leftarrow$ Mimic level + Similarity level
   //NCS Accuracy α_{High}^{Low}
4: Step-4: $FS = FG \oplus BG \oplus SP$
   //Where FS is feature set accumulated by foreground object, background object and spatial information.
5: Step-5: A α overlaying & cluttering
   $CiA\alpha FS$
   //Where CiA is complex image accuracy and impacted by FS from Step-4

Algorithm 3. Dataset Harmonizing.

Architectural bonding

The algorithm uses a deep CNN model that is employed to compare with DenseNet, VGG-19, GoogLeNet, and ResNet-50 to test the accuracy and efficiency of the presented approach in Fig. 2.

DenseNet is a CNN model that uses dense connections among layers using dense blocks. This dense block can be known as a layer. The dense block includes bottleneck layers which comprise 11 convolutions (for reducing the amount of input features maps) and 33 convolutions accordingly. Each layer gets additional inputs from all previous layers and then passes through feature maps to all consequent layers in DenseNet architecture. Every layer receives collective knowledge from all preceding layers. In Eq. (22), q^{th} layer obtains input feature maps from all preceding layers[13]:

$$b_q = I_q([b_1, b_2, ..., b_{q-1}]) \tag{22}$$

In DenseNet, every layer obtains feature maps from preceding layers, the network will be complex and thinner. For every composition layer in DenseNet, there is batch normalization (BN) and ReLU layer, then 3×3 convolutions are done using output feature maps for d channels.

The ResNet-50 model is employed to fuse with the proposed features detection and extraction to obtain maximum accurate ratios in this approach. ResNet-50 implements non-linearity at the second level. The dimensions are equal size like h and f in $y = \gamma(h, \{Mt\}) + h$. The input-output (I/O) is employed for the variations and a linear estimation M_s for the short-cut connection that executes for comparison of the dimensions in Eq. (23)[14]

$$y = f(h, \{Mt\}) + M_n\gamma \tag{23}$$

In Eq. (23), the identity mapping is assumed as a degradation problem, and M_n is applied to match the dimension. $f(h, \{Mt\})$ is used to present many convolutional layers.

VGG-19 is a CNN-based architecture with multilayered operations. VGG-19 has 16 convolutional layers to extract features at the training step[15], to transfer learning, nineteen weights layers are utilized, three layers are fully connected (FC), and an output layer at termination. Feature extraction from the input images at the initial convolutional layer, 64 kernels (3×3) are used[16]. An advanced pre-trained convolutional neural network architecture is inception. It comprises 316 layers and 350 connections. Convolution layers are 94 with various sizes where the size of the first input layer is 299×299×3. After this, batch normalization and the ReLu activation layers are supplemental. A max-pooling layer has been inserted among convolution layers. In the testing stage, categorize the images based on the softmax activation process, 10-fold cross-validation is efficient.

GoogLeNet is a deep light-weight network as its perception is effective computation and improved performance. GoogLeNet model is a relatively small amount of computational cost that is a two-phase product: firstly, optimal CNN with sparsity as introduced in[17]. Second, dimension reduction with one square convolutional layer is presented in[18]. Inception-v1 components in GoogLeNet architecture are employed for 3 filter sizes such as 1,5 and 3 squares and a maximal pooling layer. 5×5 and 3×3 filters work with 1×1 convolutional layer to reduce dimension while the maximal pooling layer is effective with 1×1 convolutional layer. GoogLeNet is used for mainly two purposes, first, to reduce the number of parameters, and second, for efficient computation. GoogLeNet model has 22 deep layers when counting layers with parameters and 27 layers when counted with pooling. 100 layers are applied in independent building blocks for GoogLeNet. Architectural bonding is defined in Algorithm 4.

1: Step-1: Generate fused DenseNet-based features [DNF]
2: Step-2: Generate ResNet-50 based fused features [RNF]
3: Step-3: Generate fused GoogLeNet based features [GNF]
4: Step-4: Generate VGG-19 based fused features [DNF]
5: Step-5: $Ab \leftarrow DNF \oplus RNF \oplus GNF \oplus DNF$
6: Step-6: $Ab \leftarrow dims$
//Where *dim* is influenced by metrics
//Where *b* is the bonding for CNN architecture crossed for multi-layers

Algorithm 4. Architectural bonding.

Using VGG-19, DenseNet, GoogLeNet, and ResNet-50 for multiple applications, the performance of computer vision is better for object detection and recognition. The feature vectors (FV) are fused with VGG-19, DenseNet, ResNet-50, and GoogLeNet; and formed feature vectors to produce powerful image signatures that are deeply signified features of shape, color, and object. The presented approach has the following contributions.

- The presented model generally collects and analyzes image contents such as color, spatial information, texture, shape, and object-created information that is formed with significant ratios of precision and recall.
- To achieve well-improved results, a novel model is presented that improves the capabilities of DenseNet, ResNet-50, VGG-19, and GoogLeNet architectures with its internal coupling.
- An image retrieval system for time-efficient, computation and storage is introduced that retrieves images in a fraction of a second.
- The presented method strengthens scale features with the bag-of-word (BoW) architecture for image retrieval, indexing, and classification.
- A new detection model and feature description are presented that accurately and effectively retrieves the relevant images related to a query from cluttered, overlay, background, foreground, and complex benchmarks.
- Introduced a new technique that achieves straddling, autocorrelation, factoring, thresholding, junctioning, regioning corner response, signature influencing, dataset harmonizing, image indexing, and deep sense image classification combined to create detail for deep improved image contents.
- Presented gray shaded images and color features depend on the image retrieval strengthens with CNNs.

- A novel methodology is proposed that effectively returns significant performance on tiny objects, complex background objects, similar textures, occluded, resized/enlarged images, cropped objects overlay ambiguous objects, color dominant arrangements, cluttered patterns, and mimicked.
- The strength of the proposed method is to retrieve the contents information for only relevant images from anchor translation instead of whole image iterations.A capsule network developed in[75] utilized attentional feature maps for classifying fire and smoke conditions. Detailed as an alternative technique for classification it allows examination of varying image classification methodologies. A presentation of our approach reveals contrasting outcomes between classification precision alongside operational effectiveness and implementation capabilities which demonstrate that multiple calculation strategies match different project specifications and dataset needs.

The proposed methodology leverages state-of-the-art CNNs for comprehensive image feature extraction and harmonization. Recent research[76] inspires the implementation of hybrid optimization that ensures efficient management of different input conditions leading to improved image feature extraction accuracy and robustness. This approach is further enriched by drawing on the principles outlined in[77]. The implementation of hybrid optimization structures this algorithm to extract useful features from heterogeneous datasets with different complexity levels thus improving model generalization abilities. The framework[78] utilizes the novel method for reliability optimization design based on rough set theory and hybrid surrogate model to manage uncertainties in data through hybrid surrogate models while performing feature extraction.

The study[79] provides inspiration for adaptive and robust algorithms that allow the extraction method to work with consistency across various image datasets. Enhanced accuracy becomes possible in complex image processing through our implementation of ensemble regression techniques described in[80]. The authors utilize causal intervention strategies that originated from[81] to enhance image feature harmonization. Features extracted from images become clearer through strategies that embed contextual understanding between media elements while boosting model performance across different imaging conditions. Similarly, the use of ResLNet in[82] establishes a specialized neural design to process lengthy complex sequences that specifically accommodate images containing extensive varieties and fine structural elements. Interpretive models in the field of modulation-based DNA storage present a useful perspective on data security and data refinement through encryption methodologies described in[83].

These principles become essential for extracting and harmonizing sensitive image data. The study[84] spotlights how feature extraction optimization becomes complicated when maintaining system performance according to the proposed methodology which assures precision across image datasets. Similarly,[85] serves as a base model to minimize distortion during feature extraction by refining the model's ability to handle different image complexities. The data integrity remains safeguarded because this method prevents essential information from disappearing throughout image harmonization. The study[86] serves as the basis to adopt its data fusion principles which merge incomplete evaluation semantics along with scheme beliefs to efficiently combine multiple image features.

The methodology is also inspired by the Flow2GNN[87] which improves GNNs' information propagation throughout the network architecture. The feature extraction process through our approach leads to effortless data retrieval from different types of datasets. Lastly, the innovative approaches presented in[88] lay the groundwork for employing AI methods in image analysis through adaptive and dynamic feature extraction techniques that support our proposed strategy. The study[89] demonstrates how domain-specific knowledge enables improved image feature extraction that adapts to multiple image conditions.

Experimentation

Datasets

The selection of an appropriate dataset is a difficult yet important task involving the classification and retrieval of images accurately. Precision proportional rates of image attributes such as cluttering, location, occlusion, quality, color, and size are important. An image retrieval system is used to find subsets that provide the corresponding database and consist of several semantic groups for different varieties. For this, experiments are performed on challenging datasets namely Zubud[19] Caltech-101[20], Corel-1000[52], Cifar-10[21], Cifar-100[21], Caltech-256[22], Corel-10000[5], COIL-100[6] and 17-Flowers[23].

Experimental results and discussion

Experiments are performed on 10 challenging benchmarks. Using the challenging measures, the efficiency, robustness, and accuracy of the presented model are computed. These measures are mean average precision (mAP), mean average recall (mAR), average retrieval precision (ARP), average recall (AR), average retrieval recall (ARR), average precision (AP), and F-measure. Accurate performance is evaluated by using recall and precision metrics. The precision is the positive estimated results and the recall is true positive ratio estimation. AP is the result for precision in the respective image category to all iterative values. ARP is the result of AP for the image category to add all categories in which every category precision value is the sum of the first category. Before plotting, the ARP results are managed to demonstrate the increment or loss progressively. Recall and precision are computed for every category[7].

$$Precision = \frac{N_{r(i)}}{N_{s(i)}} \tag{24}$$

$$Recall = \frac{N_{r(i)}}{N_q} \tag{25}$$

where $N_{r(i)}$ denotes the related images corresponding to the required image, $N_{s(i)}$ is used to recover images along with the required image and the total number of relevant images in the existing dataset[7].

$$Average\ precision(AP) = \frac{For\ every\ H_j S}{Total\ number\ of\ iteration} \tag{26}$$

$$ARP_t = \frac{For\ every\ H_j AP}{Total\ number\ of\ classes} \tag{27}$$

$$ARP = Sort(for\ each\ ARP_t) \tag{28}$$

In Eq. (26), $H_j S$ is denoted for H as a class and j is applied for the number of classes with S for the precision in the relevant class in Eq. (27). $H_j AP$ is represented H as a class and average precision described in Eq. (28). To arrange all computed values, the sort function is applied.

The AR shows recall results in the corresponding image class to all iterations. The ARP is the ratio of average recall for the image class to the sum of all classes in which each category's average recall is the sum of the first category. Before plotting, ARR ratios are arranged to demonstrate the increment or loss progressively[7].

$$Average\ recall(AR) = \frac{For\ every\ H_j Q}{Total\ number\ of\ iteration} \tag{29}$$

$$ARR_t = \frac{For\ every\ H_j AR}{Total\ number\ of\ categories} \tag{30}$$

$$ARR = Sort(for\ each\ ARR_t) \tag{31}$$

In Eq. (29), $H_j Q$ is used for H as a semantic group, and for the number of classes or semantic groups, i is represented with the Q corresponding recall in the relevant semantic group in the above Eq. (30). $H_j AR$ is used for H as class. The sort function is used to sort all calculated values in Eq. (31)[7].

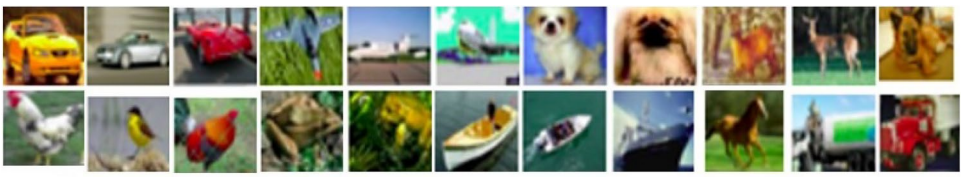
$$Moving\ average\ precision(mAP) = \frac{\mathrm{Sum(AP@H)}}{Total\ number\ of\ categories} \tag{32}$$

$$Moving\ average\ recall(mAR) = \frac{\mathrm{Sum(AR@H)}}{Total\ number\ of\ categories} \tag{33}$$

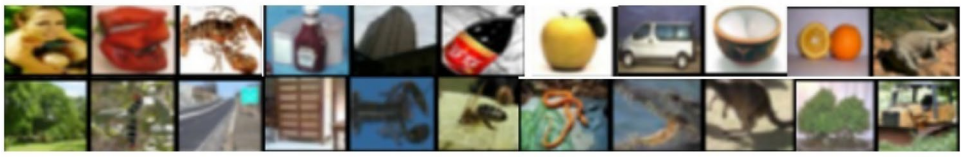
F-measure is computed as harmonic mean, multiply precision and recall results of every image class by 2 and divided by the addition of recall and the precision for each image class in Eq. (34)[49]

$$F - measure = \frac{2 \times s \times i}{s + i} \tag{34}$$

Here, s and i are denoted for the average recall and average precision rates correspondingly.

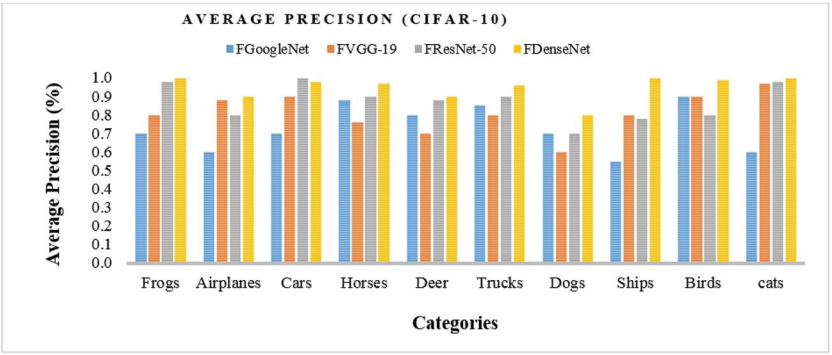


(a)

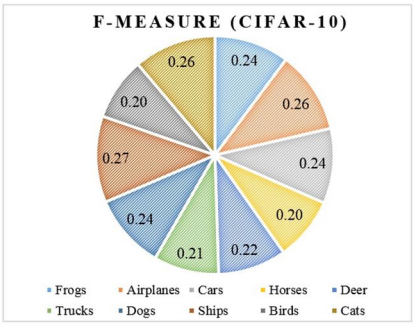


(b)

Fig. 3. Sample images from, **(a)** Cifar-10 dataset[90], and **(b)** Cifar-100 dataset[90].



(a)



(b)

Fig. 4. Results for Cifar-10dataset, **(a)** Average precision (AP) for Cifar-10 dataset, and **(b)** F-measure ratios for Cifar-10 dataset.

The retrieval time of the image taken by the presented model is 0.3 – 2.19 sec. Time differences occur in calculation due to the number of images and size of images in the database. Experiments are performed on core-i7 @2.5Ghz using 8GB RAM.

Performance on CIFAR-10 and CIFAR-100 datasets

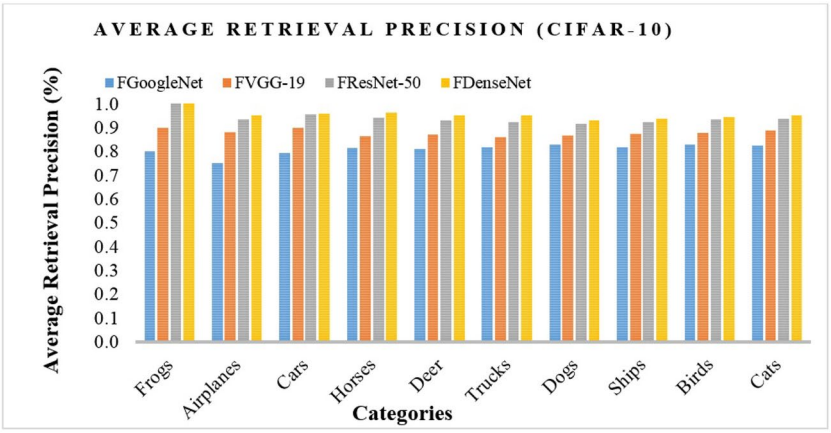
Cifar-10 dataset[90] includes several categories namely frogs, cars, birds, ships, airplanes, cats, trucks, horses, dogs, and deer. Each category contains six thousand images. In Fig. 3a, Cifar-10 dataset sample images are depicted.

Cifar-100[90] and Cifar-10 datasets are the same with 32×32 images and comprise hundred several classes. It contains different image classes namely forest, butterfly, road, lamp, palm, willow, house, bowls, mountain, rabbit, elephant, tiger, bus, clock, person, tractor, motorcycle, rocket, etc. Sample images for Cifar-100 are represented in Fig. 3b.

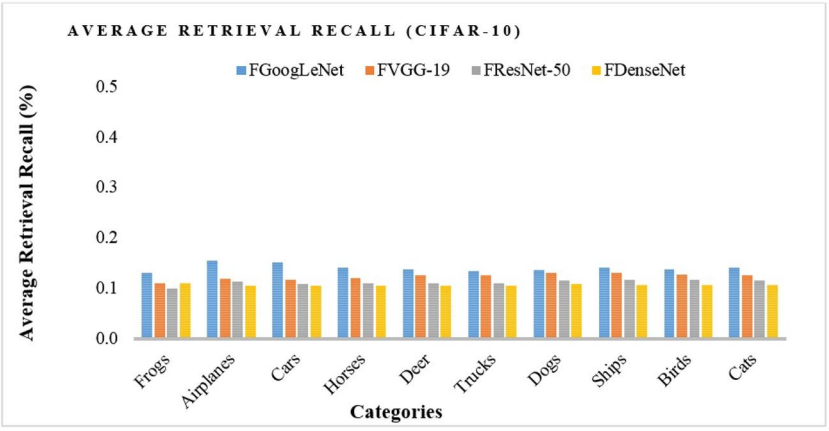
The presented algorithm uses a deep CNN model that is employed for the comparison with DenseNet, ResNet-50, GoogLeNet, and VGG-19 to test the accuracy and effectiveness of the presented model. Experiments are performed on the Cifar-10 dataset to test the affectivity and the accurateness of the presented algorithms. The fused convolutional neural networks models such as DenseNet, VGG-19, GoogleNet, and ResNet-50, are now formulated as FDenseNet, FResNet-50, FVGG-19, and FGoogleNet due to the incorporation of the presented approach with these CNN models and the presented methodology in Fig. 1 also illustrates the similar features fusion.

The presented model has outstanding average precision and F-measure results in all semantic groups for the Cifar-10 database, as shown in Fig. 4a. In Fig. 4b, the proposed algorithm shows remarkable F-measure outcomes for mimicked, tiny, cluttered, overlay, and complex background objects of the Cifar-10 dataset. Furthermore, the presented model reports significant AP ratios with FDenseNet, FResNet-50, FVGG-19, and FGoogleNet of Cifar-10. The presented model depicts remarkable AP results using FDenseNet in the image groups of Cifar-10 such as frogs, cars, airplanes, horses, trucks, birds, and cats in Fig. 5a. The outstanding precision achieved in Cifar-10 with tiny images is focused on the methodology at object recognition steps.

In Fig. 5a, the presented model provides significant ARP ratios with FDenseNet, FResNet-50, FVGG-19, and FGoogleNet of Cifar-10. In Fig. 5a, the presented algorithm depicts above 90% ARP results using FDenseNet, above 85% ARP results using FResNet-50, above 79% ARP results with FVGG-19 and 73% ARP results by FGoogLeNet for most semantic groups of Cifar-10. In Fig. 5b, the proposed model provides the highest ARP results with FDenseNet. The proposed approach also shows above 0.11 ARR results with FDenseNet, above 0.12 ARR ratios by FResNet-50, above 0.13 ARR results with FVGG-19, and 0.14 using FGoogLeNet for the Cifar-10 dataset.

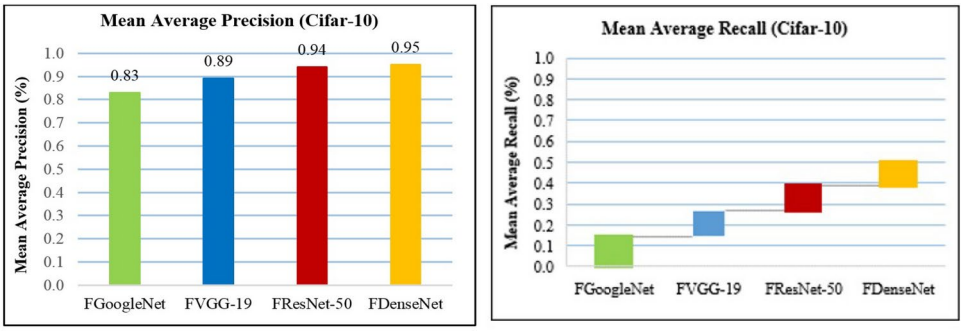


(a)



(b)

Fig. 5. Results for Cifar-10dataset, **(a)** Average retrieval precision (ARP) for Cifar-10 dataset, and **(b)** Average retrieval recall (ARR) ratios for Cifar-10 dataset.



(a) (b)

Fig. 6. Results for Cifar-10dataset, **(a)** Mean average precision (mAP) for Cifar-10 dataset, and **(b)** Mean average recall (mAR) for Cifar-10 dataset.

The presented model reports significant mAP ratios with VGG-19, GoogleNet, DenseNet, and ResNet-50 of Cifar-10. In Fig. 6a, the algorithm provides outstanding mAP 95% using DenseNet, 94% mAP ratios with ResNet-50, 89% mAP ratios with VGG-19 and 83% mAP for GoogLeNet. The proposed approach also shows above 0.11 mAR results with DenseNet, 0.12 mAR ratios by ResNet-50, 0.13 ARR results with VGG-19 and 0.14 byGoogLeNet for Cifar-10 in Fig. 6b.

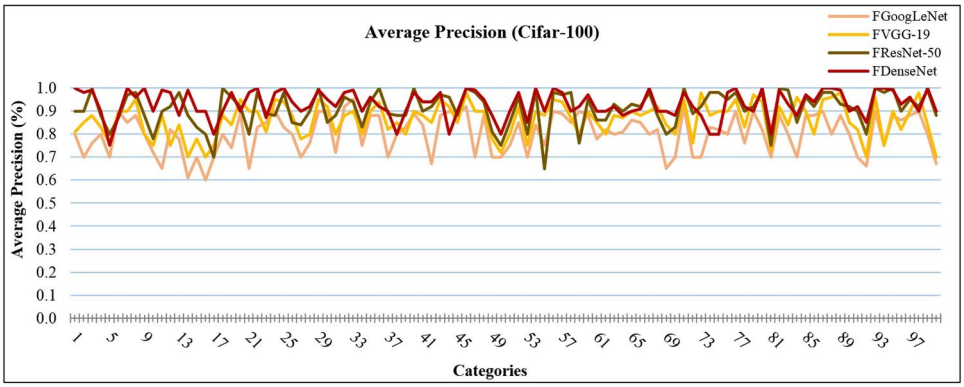
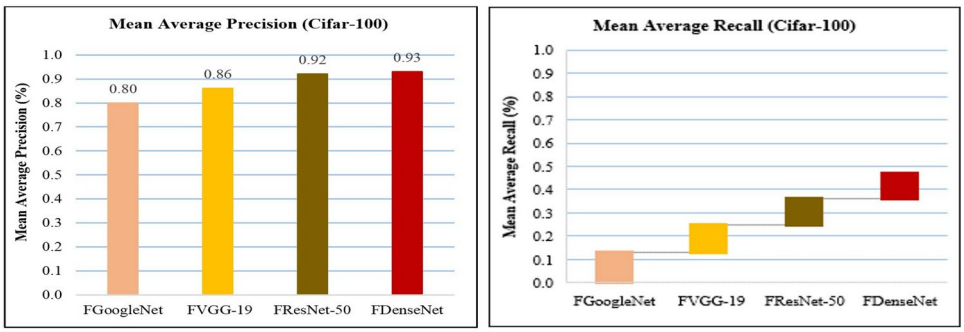


Fig. 7. Average precision (AP) for Cifar-100 dataset.



(a) (b)

Fig. 8. Results for Cifar-100 dataset, **(a)** Mean average precision (mAP) for Cifar-100 dataset, and **(b)** Mean average recall (mAR) for Cifar-100 datasett.

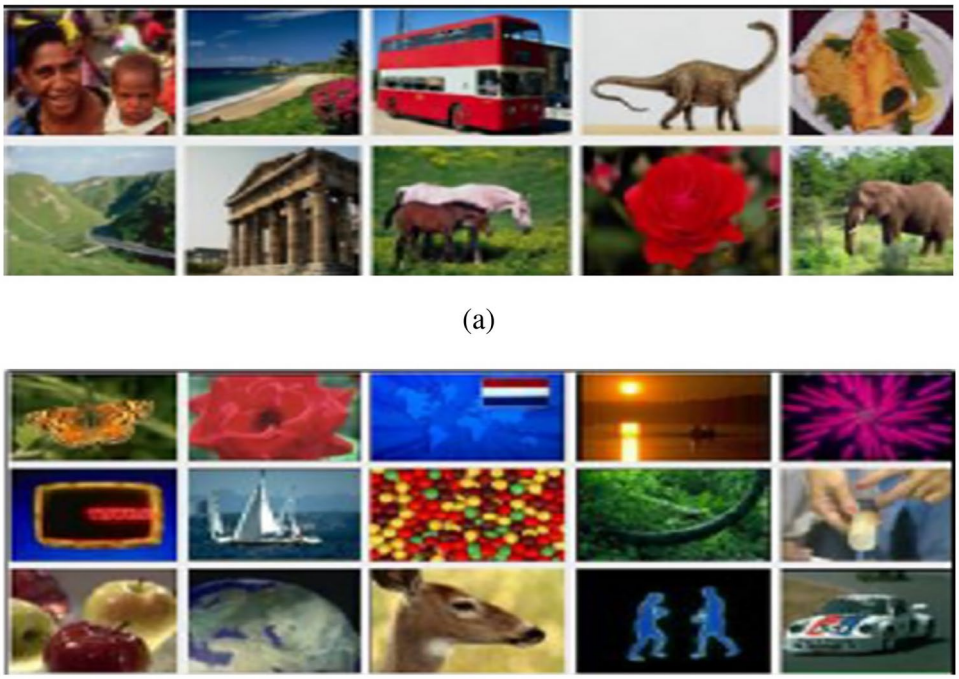
In Fig. 7, the presented model outperforms the average precision ratios of the Cifar-100 database. It is noted that the presented algorithm has outstanding results in Fig. 7 for the Cifar-100 dataset in most semantic groups. Color coefficients, distance measuring, L1 and L2 norms, color signatures, architectural bonding, signature influencing, dataset harmonizing, thresholding, functioning, factoring, and regioning are used with CNN features for effective, efficient, and accurate image indexing and classification. Several images with various image categories like horse, bear, girl, boy, chair, cloud, crocodile, rose, dolphin, bus, house, baby, train, caterpillar, butterfly, bridge, bee, lion, lizard, orchid, lobster, turtle, road, plate, trout, pear, rocket, ray, table, whale, television, telephone, woman tiger, trout, seal, tractor, shark, shrew, squirrel, spider, truck, wardrobe, wolf, willow-tree and worm with various colors and sizes are reported highest average precision results. Other image groups namely castle, camel, can, bicycle, bowl, bottle, cattle, possum, clock, streetcar, rabbit, sweet pepper, sunflower, snail, tiger, skyscraper, porcupine, raccoon, poppy, leopard, apple, otter and lizard are also precisely classified and achieved significant average precision results in most image groups of Cifar-100 dataset.

In Fig. 8a, the presented model illustrates 93% mAP results using FDenseNet, 92% mAP ratios with FResNet-50, 81% for FGoogLeNet and 86% using FVGG-19 of Cifar-100 dataset. The presented algorithm shows marvelous mean average precision ratios for the mimicked, tiny, multiple objected foreground and occupied background objects due to its efficient image recognition capability. The proposed approach outperforms complex, overlay, overlapping, and cluttered objects. In the mAR graph, the model is depicted as 0.11 mAR with FResNet-50, 0.11 with FDenseNet, 0.12 with FVGG-19, and 0.13 with FGoogLeNet of Cifar-100 dataset in Fig. 8b. In dataset harmonizing, the presented algorithm deals equally even for the large and small sizes of images. For complex, foreground and background content-based images that are not readable and objects are intermixed or vice versa.

Performance on COREL-1000 and COREL-10000 datasets

Corel-1k database is normally used to retrieve, index, and organize the images for categorization tasks. Corel-1000 database comprises several image categories namely dinosaurs, horses, mountains, flowers, beaches, buildings, foods, elephants, buses, and Africans. This dataset contains 10 groups with every comprised of a hundred selected images. Various sample images are depicted in Fig. 9a.

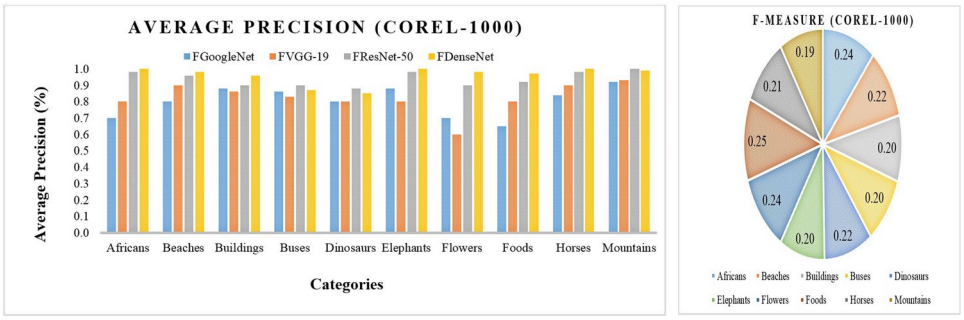
Corel-10000 dataset contains 100 classes with 100 images. The images in the Corel-10000 database are from various contents including flowers, hospitals, foods, shining stars, sunset, ketches, butterflies, human textures, texts, trees, planets, cars, animals and flags, etc. presented in Fig. 9b. The presented algorithm works at color



(a)

(b)

Fig. 9. Sample images, **(a)** Some images from Corel-1000 dataset[24]; **(b)** Some images from Corel-10k dataset[19].



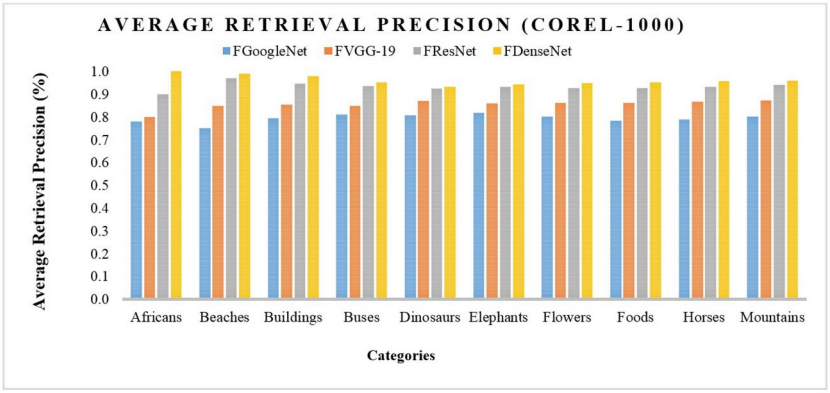
(a) (b)

Fig. 10. **(a)** Average precision (AP) for Corel-1000 dataset; **(b)** F-measure ratios for Corel-1000 dataset.

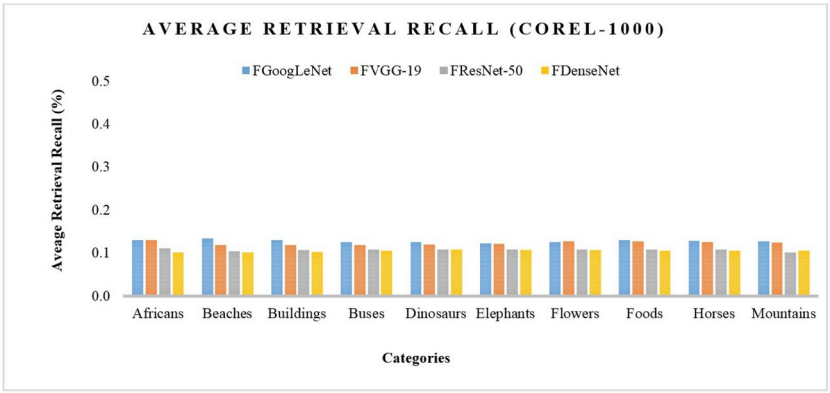
coefficients by applying appropriate displacements at L1 and L2 levels and does not cater to the explanatory aspects of feature sets by only selecting the heads that are thereby connected with the signature influencing dimension.

In Fig. 10a, the presented approach is reported with average precision results for Corel-1000. The presented model efficiently and effectively categorizes the images from several types of image groups including various foreground and background images, complex blobs, overlay, and cluttered objects due to color coefficients, distance measuring, L1 and L2 norms, color signatures, architectural bonding, signature influencing, dataset harmonizing, thresholding, factoring, and regioning methods make it possible to competently categorize the images.

The approach has marvelous average precision results for Corel-1000 reporting the dominant performance of the proposed model in most image classes due to its color channeling, autocorrelation, L2 normalization, shape parameters, and straddling. The presented algorithm depicts better performance in most of the image groups including Africans, beaches, buildings, horses, flowers, foods, and elephants using FDenseNet in Fig. 10a. The presented model also reports significant AP results by using FGoogLeNet, FVGG-19, and FResNet-50. The model presents remarkable f-measure ratios with FVGG-19, FDenseNet, FResNet-50, and FGoogLeNet of the Corel-1000 dataset in Fig. 10b. The significant precision achieved in Corel-1000 with cluttered and overlay images is focused on the dataset harmonizing algorithm at step 3 and step 4.

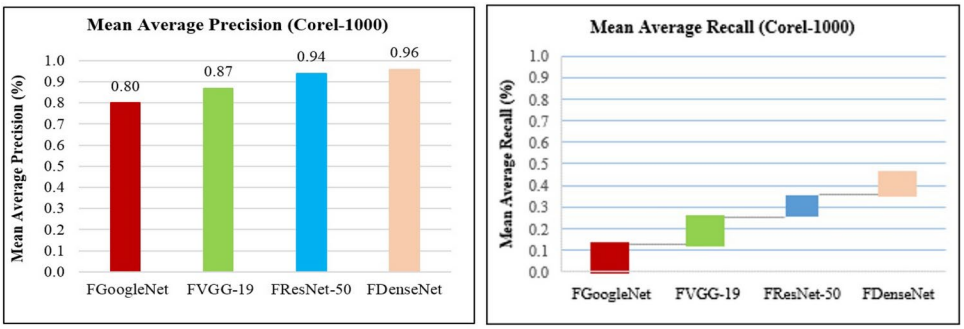


(a)



(b)

Fig. 11. **(a)** Average retrieval precision (ARP) for Corel-1000 dataset, and **(b)** Average retrieval recall (ARR) ratios for Corel-1000 dataset.



(a) (b)

Fig. 12. **(a)** Mean average precision (mAP) for Corel-1000 dataset, and **(b)** Mean average recall (mAR) for Corel-1000 dataset.

In Fig. 11a, the model provides significant ARP ratios with FDenseNet, FResNet-50, FVGG-19, and FGoogleNet of Corel-1000. The approach depicts above 95% ARP results using FDenseNet, above 93% ARP results using FResNet-50, above 88% ARP ratios with FVGG-19, and above 92% ARP results by FGoogLeNet in most semantic image groups of Corel-1000 in Fig. 11a. In Fig. 11b, the presented algorithm provides the highest ARP results with FDenseNet. The approach also shows above 0.11 ARR results with FDenseNet, above 0.11 ARR ratios by FResNet-50, above 0.12 ARR results with FVGG-19, and above 0.13 using FGoogLeNet for Corel-1000.

In Fig. 12a, the presented method illustrates 96% mAP results using FDenseNet, 94% mAP ratios with FResNet-50, 80% for FGoogLeNet, and 87% using FVGG-19 of the Corel-1000 dataset. In the mAR graph, the presented algorithm is depicted as 0.12 mAR with FResNet-50, 0.11 mAR with FDenseNet, 0.13 with FVGG-

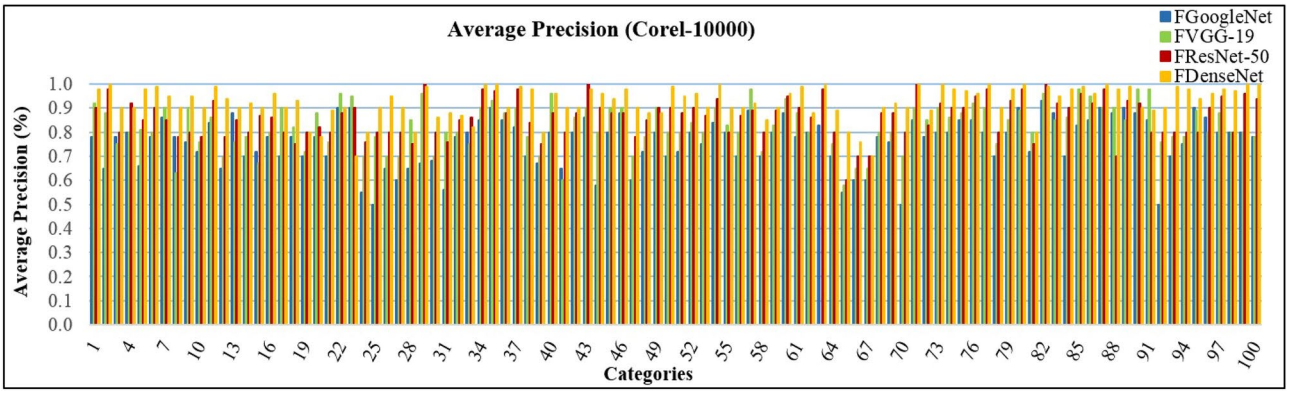
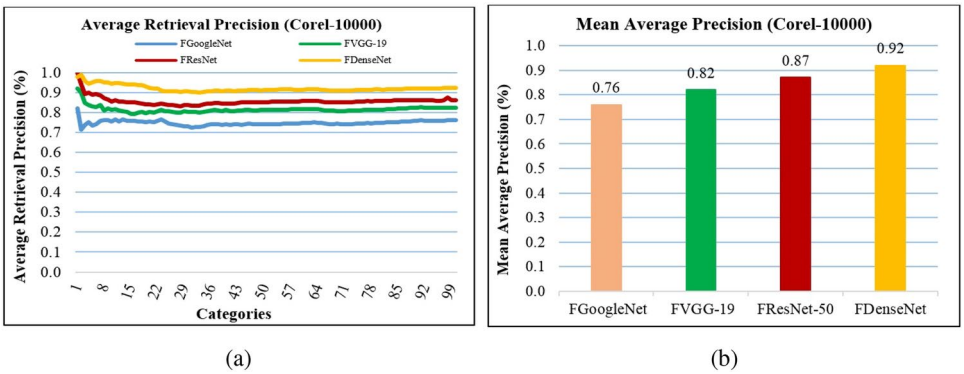


Fig. 13. Average precision (AP) for Corel-10000 dataset.



(a) (b)

Fig. 14. **(a)** Average retrieval precision for Corel-10000 dataset, and **(b)** Mean average precision (mAP) ratios for Corel-10000 dataset.

19, and 0.14 with FGoogLeNet of the Corel-1000 dataset in Fig. 12b. The recent research contribution presents feature fusions that finally provide large feature-set sizes that directly influence the experimental cost and there they lose values in merging. However, in the proposed algorithm and proposed methodology, the signatures are not influenced due to signature length. These are manipulated in such a way that the computational cost is minimal. This research first time introduces diverse CNN complex architectures that have never been used in experimentation before collectively by any researcher. These datasets and CNNs are incorporated with the presented algorithm. This is a heterogeneous architectural bonding that is presented in this research and shows the high level of precision for complex datasets with deep image sensing.

In Fig. 13, the presented model has reported significant average precision ratios of the Corel-10000 database. In Fig. 13, 1t is observed that the Corel-10000 dataset has outstanding results in most of the image groups. Color coefficients, distance measuring, normalization, color signatures, architectural bonding, signature influencing, dataset harmonizing, thresholding, junctioning, factoring, and regioning are used with CNN features for effective, efficient, and accurate image indexing and classification. The presented algorithm reported the highest average precision results with FDenseNet, FResNet-50, FGoogLeNet, and FVGG-19 due to color channeling, straddling, anisotroping, and autocorrelation. In dataset harmonizing, the presented algorithm deals equally even for the large and small sizes of images. For complex, foreground and background content-based images that are not readable and objects are intermixed or vice versa. The presented algorithm extracts a similar level of precision. The significant precision achieved in Corel-10000 with cluttered and complex images is focused on the architectural bonding algorithm at step 4, step 6, and step 7.

In Fig. 14b, the presented method illustrates 92% mAP results using FDenseNet, 87% mAP ratios with FResNet-50, 76% for FGoogLeNet, and 81% using FVGG-19 of Corel-10000 dataset. The approach is depicted above 90% mAP with FDenseNet, above 85% mAP using FResNet-50, above 80% mAP with FVGG-19, and above 75% mAP results with FGoogLeNet of Corel-10000 dataset in Fig. 14a. Corel-10000 dataset contains different shapes and textural images using the same patterns, shapes, and colors whereas the other image groups comprise different object patterns. The mean average precision results of Corel-10000 show the robustness of the presented approach.

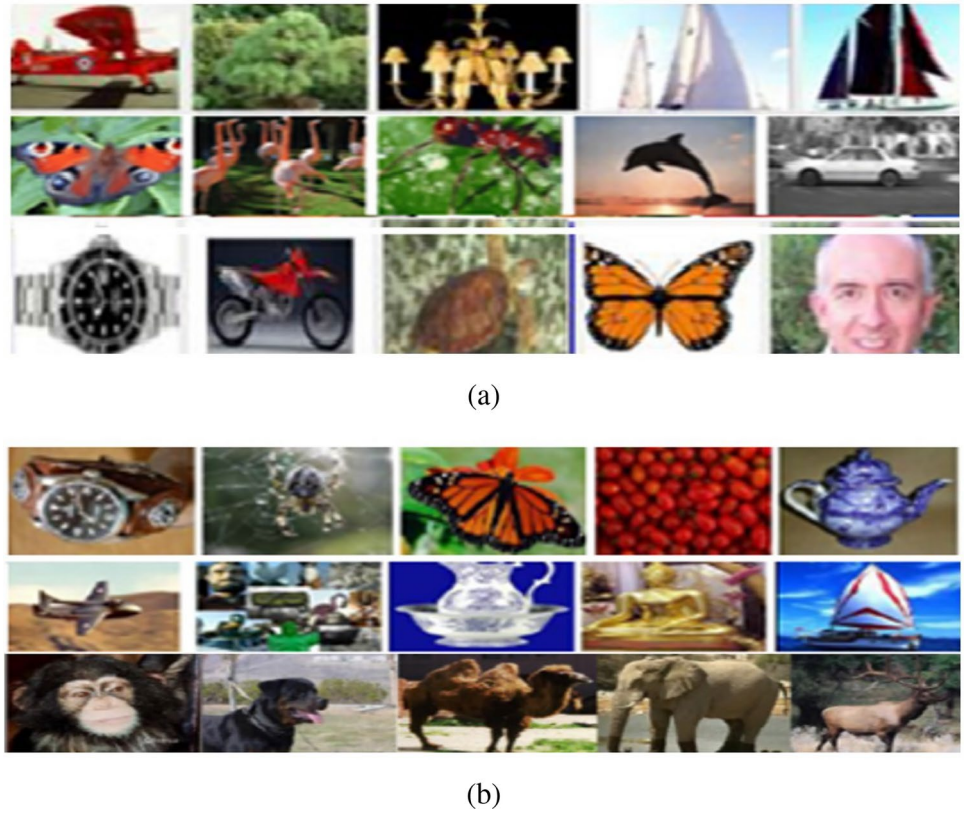


Fig. 15. Sample images from **(a)** Caltech-101 dataset, and **(b)** Caltech-256 dataset[91].

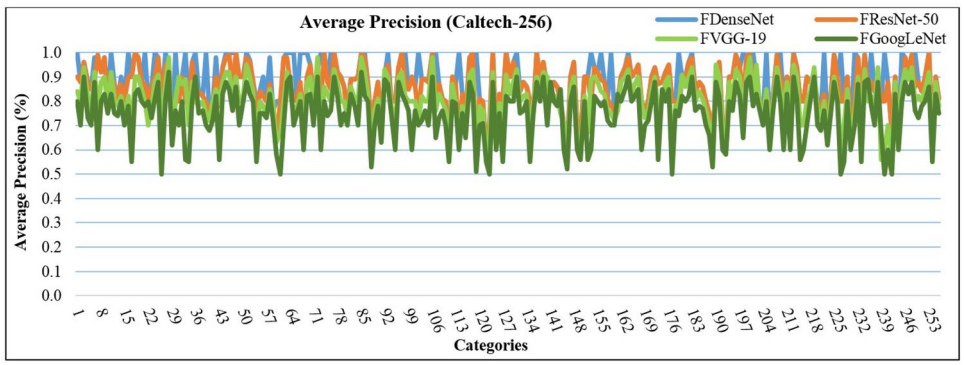


Fig. 16. Average precision (AP) for Caltech-256 dataset.

Performance on CALTECH-101 and CALTECH-256 datasets

Caltech-101[20] database is used to classify and retrieve the images, match the objects, and recognize the images illustrated in Fig. 15a. Some images for Caltech-101 are illustrated in Fig. 15a. This dataset includes more than 9000 images. Caltech-101 database comprises several semantic groups including tortoise, bonsai, wristwatch, airplane, leopard, chandeliers, face-easy, brain, motorbikes, things, buddha, ketch, butterfly, face, ewer, etc. These semantic groups are used to test the robustness of the presented algorithm which is accomplished by sharing the rounded, spatial values and multi-shaped objects with texture information comprised with the color coefficients.

Caltech-256[22] comprises over thirty thousand images. This database covers different semantic groups such as back-pack, bonsai, bulldozer, tomato, spider, wristwatch, cactus, teapot, teddy bear, billiards, boxing gloves, swan, tree, and butterfly airplane. In this dataset, all classes are very important due to textural patterns, foreground objects, and cluttered objects. Some sample images for Caltech-256 are illustrated in Fig. 15b.

Figure 16 shows outstanding AP results for the Caltech-256 dataset. It is also observed that the algorithm reports better AP results in most semantic classes for Caltech-256 having different colors, textures, and shapes. The proposed approach applies color channeling, thresholding, distance measuring, and straddling by CNN features for efficient image indexing and classification. Several images with various image groups such as

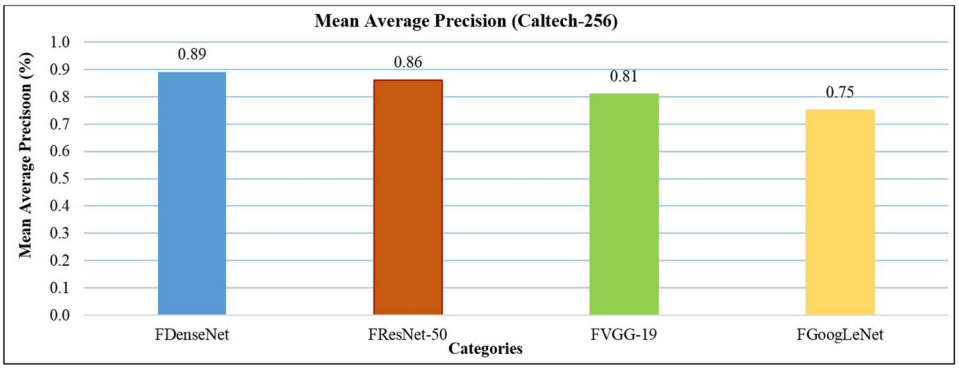


Fig. 17. Mean average precision (mAP) for Caltech-256 dataset.

animals, electronics, foods, transportation, plants, sports, and flowers with various colors, textures, and shapes are used. The proposed approach reports the highest AP results above 90% with FDenseNet in most of the image groups for Caltech-256. The presented algorithm shows above 87% AP results with FResNet-50, above 80% AP ratios by FVGG-19, and above 76% AP ratios by FGoogLeNet in most semantic groups. Other semantic groups of Caltech-256 namely humans, games, fun, apparel, musical, weapons, households, structures, tools, insects, religious and natural are classified efficiently and achieved the highest accuracy in most of the image groups of Caltech-256. The presented algorithm extracts a similar level of precision. The significant precision achieved in Caltech-256 with cluttered and complex images is focused on the signature influencing algorithm at step 3 and step 4.

In Fig. 17, the method reports 89% mAP results using FDenseNet, 86% mAP ratios with FResNet-50, 75% for FGoogLeNet and 81% using FVGG-19 of Caltech-256 dataset. Different semantic groups are capable of classifying with CNN features using the color signature, color coefficients, distance measuring, straddling, peaked, and autocorrelation in the proposed approach. The presented architectural bonding and dataset harmonizing algorithms are applied to attain the highest mean average precision results. Caltech-256 dataset contains different shapes and textures images of the same patterns whereas the other image categories comprise various object patterns. The mean average precision results of Caltech-256 show the robustness of the presented approach.

Table 1 illustrates marvelous average precision, ARR, and F-measure results for the Caltech-101 dataset. The average precision rates of the algorithm are highest in most of the image categories using FDenseNet and RFesNet-50. The proposed method provides above 95% AP ratios in most of the image groups of Caltech-101. Few image groups present 100% average precision results using FDenseNet including anchor, brain, cougar_face, garfield, lamp, revolver, water_lily, Windsor_chair, and yin_yang. The presented algorithm applies color coefficients, distance measuring, thresholding, architectural bonding, signature influencing, and dataset harmonizing with CNN features to classify images efficiently. The presented approach shows more than 87% average precision results by FVGG-19 and more than 82% average precision results by FGoogLeNet in most semantic groups for Caltech-101. The presented technique reports less than 80% average precision ratios in a few image groups due to overlay, large, mimicked, cluttered, and complex objects. Furthermore, the approach provides the highest 90% mAP ratios with FDenseNet. The presented model reports 87% mAP results with FResNet-50, 83% mAP results with FVGG-19, and 77% mAP results by FGoogLeNet. The approach outperforms with FDenseNet.

Furthermore, the technique provides the highest f-measure ratios in Table 1 for the Caltech-101 dataset. The proposed model provides f-measure rates between 18% to 27% ratios with FDenseNet, 18% to 27% ratios by FResNet-50, 18% to 26% ratios by FVGG-19 and 19% to 29% ratios by FGoogLeNet for shape dominant, cluttered, overlay, large, complex, edges dominant, color dominant objects of Caltech-101 dataset. The presented algorithm shows significant ARR results for Caltech-101 with FDenseNet, FResNet-0, FGoogLeNet, and FVGG-19 in Table 1. Hence, the part of presented algorithms presented in the fourth presented algorithm is incorporated at lower and higher levels by producing dataset harmonizing at 32×32 images of Cifar-10 and Cifar-100 as well as image size of Caltech-101. This novel algorithmic script produces better object recognition and primitive feature detection at foreground and background levels with equal correspondence for tiny and large images; which ultimately proves the dataset harmonizing concept in CBIR for the first time. In dataset harmonizing, the presented algorithm deals equally even for the large and small sizes of images. For complex, foreground and background content-based images that are not readable and objects are intermixed or vice versa. The presented algorithm extracts a similar level of precision.

Performance on ZUBUD and COIL-100 datasets

Zurich Building Database (Zubud) dataset[19] includes colored images of 201 buildings images Zurich city. These images for buildings are accessible with five images achieved at five random viewpoints. The partially complex and cluttered backgrounds are clearly present (people, trees, trams, skies, cars, etc.) with scale and orientation changes due to the photographer's position. In Fig. 18a, some Zubud dataset images are shown.

	FDenseNet			FResNet			FVGG-19			FGoogleNet		
Category	AP	ARR	F	AP	ARR	F	AP	ARR	F	AP	ARR	F
accordion	0.98	0.10	0.18	1.00	0.10	0.18	0.90	0.11	0.20	0.70	0.14	0.24
airplanes	0.99	0.10	0.18	1.00	0.10	0.18	0.87	0.11	0.20	0.80	0.13	0.22
anchor	1.00	0.10	0.18	0.98	0.10	0.18	0.90	0.11	0.20	0.85	0.12	0.21
ant	0.90	0.10	0.20	0.75	0.12	0.23	0.70	0.13	0.24	0.75	0.13	0.23
Background _Google	0.96	0.10	0.19	0.95	0.12	0.19	0.88	0.13	0.20	0.82	0.13	0.21
barrel	0.88	0.11	0.20	0.92	0.11	0.19	0.85	0.12	0.21	0.86	0.12	0.20
bass	0.90	0.11	0.20	0.94	0.11	0.19	0.90	0.11	0.20	0.81	0.12	0.21
beaver	0.95	0.11	0.19	0.88	0.11	0.20	0.80	0.12	0.22	0.70	0.13	0.24
binocular	0.80	0.11	0.22	0.65	0.13	0.25	0.75	0.13	0.23	0.65	0.15	0.25
bonsai	0.60	0.11	0.26	0.80	0.14	0.22	0.76	0.13	0.22	0.70	0.15	0.24
brain	1.00	0.11	0.18	0.98	0.11	0.18	0.90	0.12	0.20	0.80	0.13	0.22
brontosaurus	0.80	0.11	0.22	0.72	0.12	0.23	0.68	0.13	0.24	0.65	0.14	0.25
buddha	0.94	0.11	0.19	0.90	0.13	0.20	0.91	0.13	0.20	0.85	0.14	0.21
butterfly	0.85	0.11	0.21	0.80	0.12	0.22	0.78	0.12	0.22	0.70	0.13	0.24
camera	0.92	0.11	0.19	0.89	0.12	0.20	0.80	0.13	0.22	0.93	0.13	0.19
cannon	0.88	0.11	0.20	0.90	0.11	0.20	0.98	0.11	0.18	0.78	0.12	0.22
car_side	0.96	0.11	0.19	1.00	0.11	0.18	0.95	0.10	0.19	0.88	0.12	0.20
ceiling_fan	0.75	0.11	0.23	0.70	0.12	0.24	0.65	0.13	0.25	0.50	0.16	0.29
cellphone	0.88	0.11	0.20	0.86	0.13	0.20	0.80	0.14	0.22	0.70	0.17	0.24
chair	0.92	0.11	0.19	0.89	0.11	0.20	0.80	0.13	0.22	0.76	0.14	0.22
chandelier	0.93	0.11	0.19	0.98	0.11	0.18	0.94	0.12	0.19	0.88	0.12	0.20
cougar_face	1.00	0.11	0.18	0.94	0.10	0.19	0.96	0.11	0.19	0.90	0.11	0.20
crab	0.98	0.11	0.18	0.94	0.11	0.19	0.90	0.11	0.20	0.83	0.12	0.21
crayfish	0.90	0.11	0.20	0.88	0.11	0.20	0.70	0.13	0.24	0.65	0.14	0.25
crocodile	0.88	0.11	0.20	0.76	0.12	0.22	0.70	0.14	0.24	0.62	0.16	0.26
crocodile_head	0.67	0.11	0.24	0.75	0.13	0.23	0.65	0.15	0.25	0.55	0.17	0.27
cup	0.98	0.11	0.18	0.97	0.12	0.19	0.92	0.13	0.19	0.84	0.15	0.21
dalmatian	0.90	0.11	0.20	0.85	0.11	0.21	0.90	0.11	0.20	0.80	0.12	0.22
dollar_bill	0.98	0.11	0.18	0.95	0.11	0.19	0.88	0.11	0.20	0.80	0.13	0.22
dolphin	0.69	0.11	0.24	0.55	0.14	0.27	0.58	0.14	0.27	0.52	0.16	0.28
dragonfly	0.92	0.11	0.19	0.93	0.14	0.19	0.90	0.14	0.20	0.86	0.15	0.20
electric_ guitar	0.92	0.11	0.19	0.90	0.11	0.20	0.82	0.12	0.21	0.90	0.11	0.20
elephant	0.85	0.11	0.21	0.83	0.12	0.21	0.77	0.13	0.22	0.70	0.13	0.24
emu	0.90	0.11	0.20	0.89	0.12	0.20	0.80	0.13	0.22	0.76	0.14	0.22
euphonium	0.80	0.11	0.22	0.75	0.12	0.23	0.70	0.13	0.24	0.69	0.14	0.24
ewer	0.99	0.11	0.18	1.00	0.12	0.18	0.94	0.12	0.19	0.90	0.13	0.20
faces	0.94	0.11	0.19	0.98	0.10	0.18	0.90	0.11	0.20	0.82	0.12	0.21
faces_easy	0.99	0.11	0.18	1.00	0.10	0.18	0.89	0.11	0.20	0.80	0.12	0.22
ferry	0.80	0.11	0.22	0.75	0.12	0.23	0.70	0.13	0.24	0.62	0.14	0.26
flamingo	0.98	0.11	0.18	0.90	0.12	0.20	0.95	0.12	0.19	0.80	0.14	0.22
flamingo_head	0.95	0.11	0.19	0.89	0.11	0.20	0.82	0.11	0.21	0.78	0.13	0.22
garfield	1.00	0.11	0.18	0.98	0.11	0.18	0.90	0.12	0.20	0.80	0.13	0.22
gerenuk	0.79	0.11	0.22	0.60	0.13	0.26	0.60	0.14	0.26	0.55	0.15	0.27
gramophone	0.88	0.11	0.20	0.78	0.15	0.22	0.70	0.15	0.24	0.52	0.19	0.28
grand_piano	0.98	0.11	0.18	0.95	0.12	0.19	0.94	0.12	0.19	0.86	0.15	0.20
hawksbill	0.86	0.11	0.20	0.75	0.12	0.23	0.65	0.13	0.25	0.60	0.14	0.26
headphone	0.90	0.11	0.20	0.88	0.12	0.20	0.80	0.14	0.22	0.90	0.14	0.20
hedgehog	0.98	0.11	0.18	1.00	0.11	0.18	0.94	0.12	0.19	0.88	0.11	0.20
helicopter	0.88	0.11	0.20	0.84	0.11	0.21	0.80	0.12	0.22	0.72	0.13	0.23
ibis	0.80	0.11	0.22	0.75	0.13	0.23	0.70	0.13	0.24	0.60	0.15	0.26
inline_skate	0.97	0.11	0.19	0.90	0.12	0.20	0.80	0.13	0.22	0.78	0.15	0.22
joshua_tree	0.96	0.11	0.19	0.95	0.11	0.19	0.90	0.12	0.20	0.84	0.12	0.21
kangaroo	0.79	0.11	0.22	0.65	0.13	0.25	0.60	0.14	0.26	0.68	0.13	0.24
ketch	0.82	0.11	0.21	0.85	0.14	0.21	0.80	0.15	0.22	0.73	0.14	0.23
Continued												

Category	FDenseNet			FResNet			FVGG-19			FGoogleNet		
	AP	ARR	F	AP	ARR	F	AP	ARR	F	AP	ARR	F
lamp	1.00	0.11	0.18	0.98	0.11	0.18	0.82	0.12	0.21	0.80	0.13	0.22
laptop	0.80	0.11	0.22	0.75	0.12	0.23	0.70	0.13	0.24	0.65	0.14	0.25
leopard	0.98	0.11	0.18	1.00	0.12	0.18	0.91	0.13	0.20	0.85	0.14	0.21
llama	0.99	0.11	0.18	0.98	0.10	0.18	0.90	0.11	0.20	0.83	0.12	0.21
lobster	0.85	0.11	0.21	0.83	0.11	0.21	0.80	0.12	0.22	0.74	0.13	0.23
lotus	0.90	0.11	0.20	0.86	0.12	0.20	1.00	0.11	0.18	0.82	0.13	0.21
mandolin	0.98	0.11	0.18	0.95	0.11	0.19	0.90	0.11	0.20	0.80	0.12	0.22
mayfly	0.98	0.11	0.18	0.95	0.11	0.19	0.93	0.11	0.19	0.88	0.12	0.20
menorah	0.95	0.11	0.19	0.92	0.11	0.19	0.95	0.11	0.19	0.87	0.11	0.20
metronome	0.99	0.11	0.18	0.97	0.11	0.19	0.89	0.11	0.20	0.88	0.11	0.20
minaret	0.95	0.11	0.19	0.88	0.11	0.20	0.70	0.13	0.24	0.67	0.13	0.24
motorbikes	0.95	0.11	0.19	1.00	0.11	0.18	0.92	0.13	0.19	0.88	0.13	0.20
nautilus	0.80	0.11	0.22	0.85	0.11	0.21	0.83	0.11	0.21	0.78	0.12	0.22
octopus	0.86	0.11	0.20	0.89	0.12	0.20	0.82	0.12	0.21	0.77	0.13	0.22
okapi	0.85	0.11	0.21	0.90	0.11	0.20	0.80	0.12	0.22	0.70	0.14	0.24
pagoda	0.66	0.11	0.25	0.70	0.13	0.24	0.70	0.13	0.24	0.65	0.15	0.25
panda	0.98	0.11	0.18	0.90	0.13	0.20	0.88	0.13	0.20	0.84	0.14	0.21
pigeon	0.88	0.11	0.20	0.83	0.12	0.21	0.77	0.12	0.22	0.73	0.13	0.23
pizza	0.80	0.11	0.22	0.75	0.13	0.23	0.70	0.14	0.24	0.66	0.14	0.25
platypus	0.94	0.11	0.19	0.90	0.12	0.20	0.83	0.13	0.21	0.79	0.14	0.22
pyramid	0.89	0.11	0.20	0.83	0.12	0.21	0.79	0.12	0.22	0.72	0.13	0.23
revolver	1.00	0.11	0.18	0.98	0.11	0.18	0.92	0.12	0.19	0.88	0.13	0.20
rhino	0.60	0.11	0.26	0.55	0.14	0.27	0.90	0.11	0.20	0.82	0.12	0.21
rooster	0.93	0.11	0.19	0.90	0.15	0.20	0.85	0.11	0.21	0.80	0.12	0.22
saxophone	0.88	0.11	0.20	0.78	0.12	0.22	0.67	0.13	0.24	0.60	0.15	0.26
schooner	0.78	0.11	0.22	0.75	0.13	0.23	0.80	0.14	0.22	0.70	0.15	0.24
scissors	0.98	0.11	0.18	1.00	0.12	0.18	0.93	0.12	0.19	0.88	0.13	0.20
scorpion	0.90	0.11	0.20	0.93	0.10	0.19	0.90	0.11	0.20	0.95	0.11	0.19
sea_horse	0.98	0.11	0.18	0.91	0.11	0.20	0.85	0.11	0.21	0.78	0.12	0.22
snoopy	0.85	0.11	0.21	0.78	0.12	0.22	0.75	0.13	0.23	0.75	0.13	0.23
soccer_ball	0.95	0.11	0.19	0.90	0.12	0.20	0.83	0.13	0.21	0.73	0.14	0.23
stapler	0.90	0.11	0.20	0.89	0.11	0.20	0.80	0.12	0.22	0.74	0.14	0.23
starfish	0.60	0.11	0.26	0.60	0.14	0.26	0.63	0.14	0.25	0.58	0.15	0.27
stegosaurus	0.98	0.11	0.18	0.93	0.14	0.19	0.90	0.13	0.20	0.85	0.15	0.21
stop_sign	0.94	0.11	0.19	1.00	0.10	0.18	0.95	0.11	0.19	0.90	0.11	0.20
strawberry	0.80	0.11	0.22	0.65	0.13	0.25	0.60	0.14	0.26	0.55	0.15	0.27
sunflower	0.85	0.11	0.21	0.90	0.13	0.20	0.98	0.13	0.18	0.82	0.15	0.21
tick	0.76	0.11	0.22	0.80	0.12	0.22	0.74	0.12	0.23	0.70	0.13	0.24
trilobite	0.90	0.11	0.20	0.85	0.12	0.21	0.78	0.13	0.22	0.72	0.14	0.23
umbrella	0.99	0.11	0.18	0.90	0.11	0.20	0.88	0.12	0.20	0.91	0.12	0.20
watch	0.96	0.11	0.19	0.95	0.11	0.19	0.90	0.11	0.20	0.83	0.12	0.21
water_lilly	1.00	0.11	0.18	0.98	0.10	0.18	0.94	0.11	0.19	0.90	0.12	0.20
wheelchair	0.95	0.11	0.19	0.87	0.11	0.20	0.75	0.12	0.23	0.70	0.13	0.24
wild_cat	0.90	0.11	0.20	0.88	0.11	0.20	0.84	0.13	0.21	0.78	0.14	0.22
windsor _chair	1.00	0.11	0.18	0.98	0.11	0.18	0.90	0.12	0.20	0.80	0.13	0.22
wrench	0.95	0.11	0.19	0.89	0.11	0.20	0.90	0.11	0.20	0.83	0.12	0.21
yin_yang	1.00	0.11	0.18	0.99	0.11	0.18	0.98	0.11	0.18	0.92	0.11	0.19

Table 1. AP, ARR and F-measure ratios for Caltech-101 dataset.

The dataset of the Columbia object image library (COIL-100) comprises 100 semantic groups. Every category contains seventy-two images. This includes various categories including tomato, white cup, rolaids, herb-box, mud pot, pink cup, car, statue, soft edri, cat, truck, fancy feast stick, frog, and jug. Figure 18b shows the sample images of the Columbia object image library (COIL-100) database.

Figure 19 shows outstanding AP results for the Zubud dataset. It is also noted that the algorithm reports improved AP results in most of the image classes of the Zubud dataset having different colors, textures, and

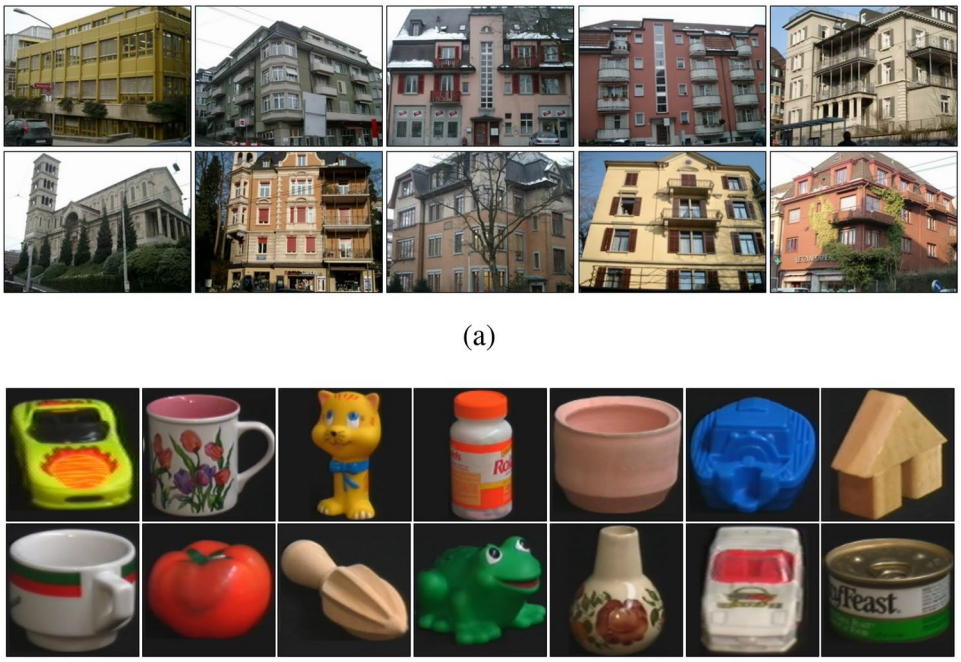


Fig. 18. Samples images from **(a)** Zubud dataset[19], and **(b)** COIL-100 dataset[92].

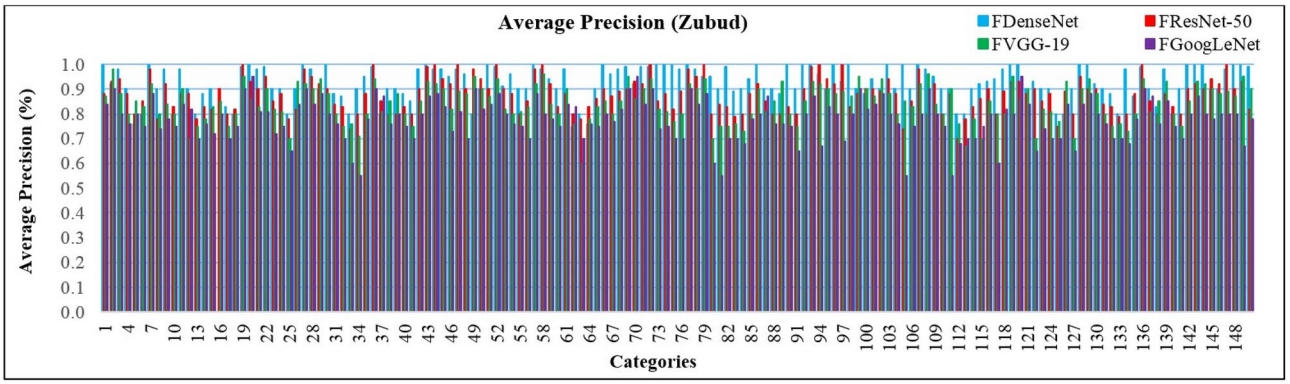


Fig. 19. Average precision (AP) for Zubud dataset.

shapes. The proposed approach applies color channeling, thresholding, distance measuring, and straddling by CNN features for efficient image indexing and classification. The proposed approach reports the highest AP results above 92% with FDenseNet in most of the image groups for Zubud. This proposed algorithm shows above 88% AP results with FResNet-50, above 81% AP results by FVGG-19, and above 77% AP results with FGoogLeNet in most of the semantic groups of Zubud.

The presented model shows F-measure results between 18% to 23% ratios with FDenseNet, 18% to 24% ratios by FResNet-50, 19% to 24% ratios by FVGG-19 and 19% to 27% rates by FGoogLeNet for large, cluttered, overlay, complex, color dominant objects of Zubud dataset in Figs. 20 and 21.

In Fig. 22a, the approach depicts 93% mAP results using FDenseNet, 88% mAP ratios with FResNet-50, 78% for FGoogLeNet and 84% using FVGG-19 of Zubud dataset. It is noted that the presented algorithm shows increased mAP results for the Zubud dataset with different shapes, colors, patterns, and textures. Straddling, anchor points, autocorrelation, shape parameters, factoring, regioning, thresholding and junctioning, and color coefficients with CNN features make this possible to efficiently classify and retrieve images efficiently. Hence, the part of presented algorithms presented in the fourth presented algorithm is incorporated at lower and higher levels by producing dataset harmonizing of Zubud. This novel algorithmic script produces better object recognition and primitive feature detection at foreground and background levels with equal correspondence for tiny and large images; which ultimately proves the dataset harmonizing concept in CBIR for the first time. In the

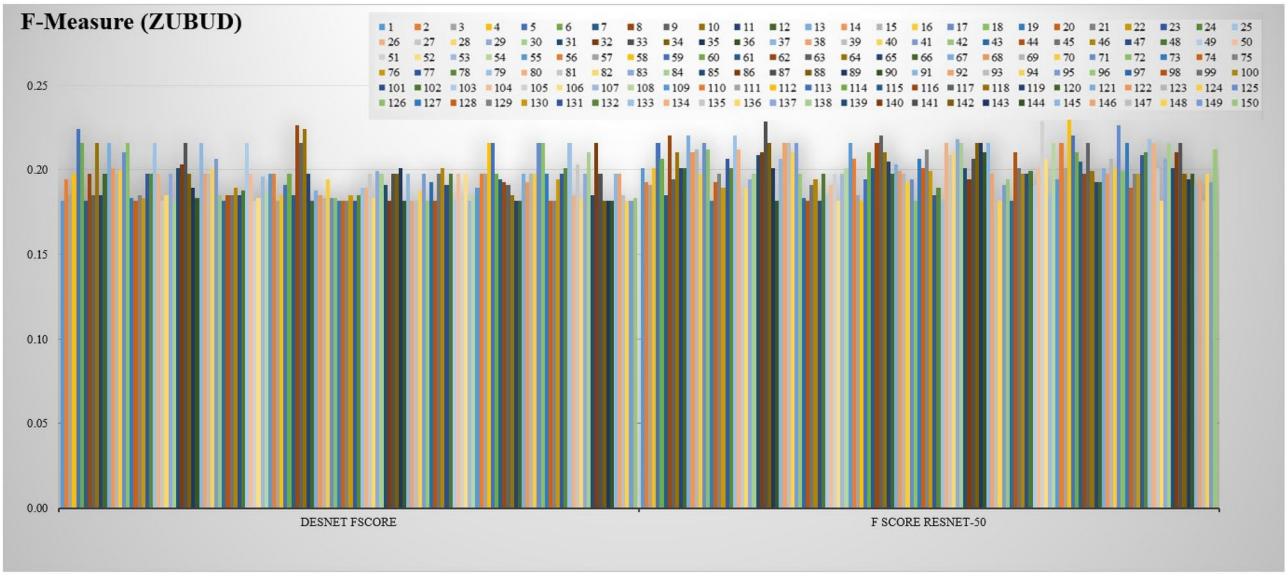


Fig. 20. F-measure ratios for Zubud dataset.

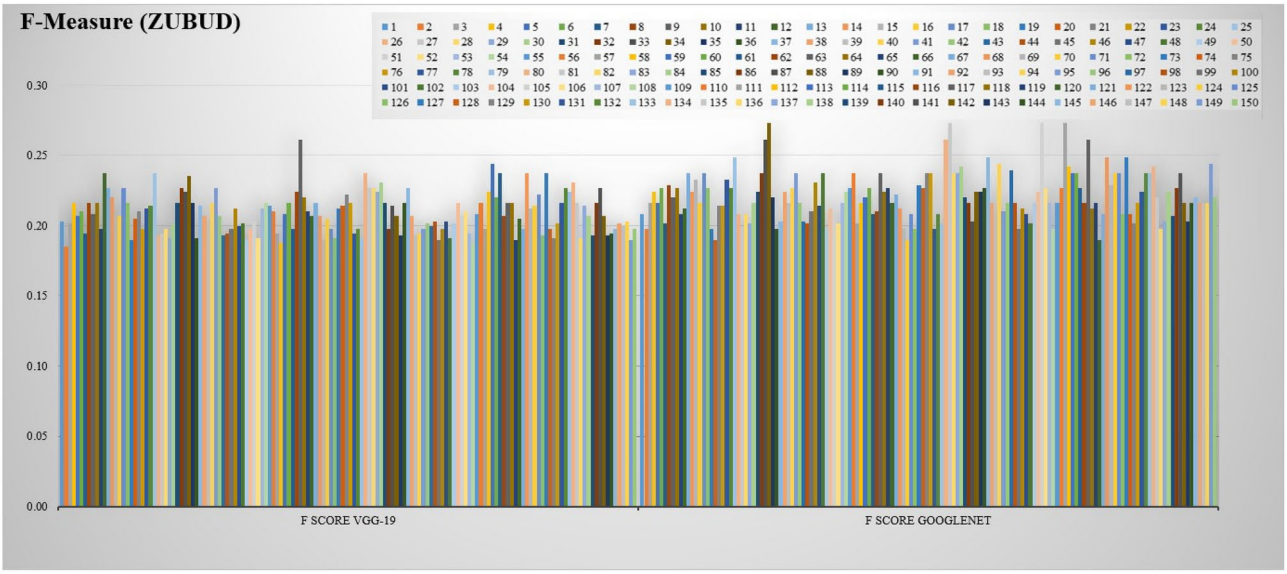
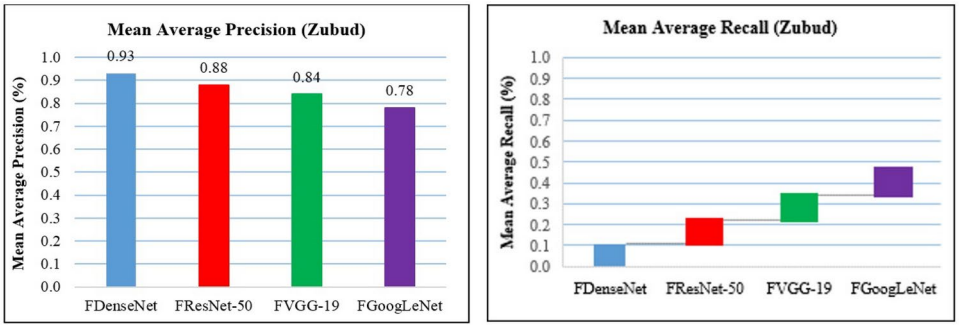


Fig. 21. F-measure ratios for Zubud dataset.

mAR graph, the presented model is depicted as 0.12 with FResNet-50, 0.11 with FDenseNet, 0.13 with FVGG-19, and 0.14 with FGoogLeNet of Zubud dataset in Fig. 22b.

The presented algorithm outperforms AP ratios for COIL-100 in Fig. 23. It is also noted that the research approach reports improved AP results in most of the image classes of the COIL-100 dataset having different colors, textures, and shapes. The proposed approach applies color channeling, thresholding, distance measuring, and straddling by CNN features for efficient image indexing and classification. The approach reports the highest AP results above 93% with FDenseNet in most of the image groups for COIL-100. This proposed model shows above 89% AP results with FResNet-50, above 83% AP results by FVGG-19, and above 80% AP results with FGoogLeNet in most semantic groups of COIL-100.

The proposed model shows f-measure results between 18% to 26% ratios with FDenseNet, 18% to 24% ratios by FResNet-50, 18 to 27% results by FVGG-19 and 19% to 24% ratios by FGoogLeNet for cluttered, overlay, large, edges dominant, complex, and color dominant objects of COIL-100 dataset in Figs. 24 and 25. In Fig. 26, the presented algorithm provides significant ARP ratios with FVGG-19, FDenseNet, FGoogleNet, and FResNet-50 of the COIL-100 dataset. In Fig. 26, the approach depicts above 95% ARP results using FDenseNet, above 93% ARP results using FResNet-50, above 88% ARP results by FVGG-19 and above 92% ARP results by FGoogLeNet for most semantic groups of COIL-100.



(a) (b)

Fig. 22. **(a)** Mean average precision (mAP) for the Zubud dataset, and **(b)** Mean average recall (mAR) for the Zubud dataset.

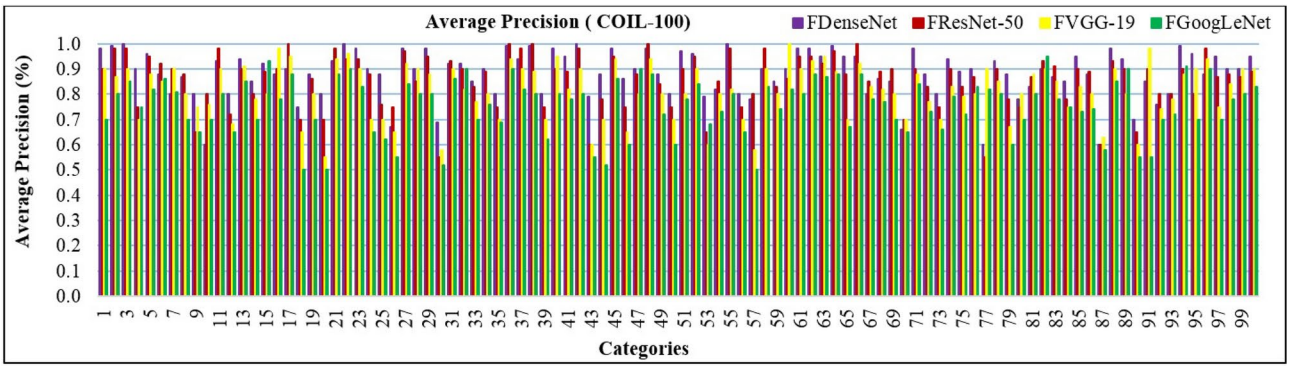


Fig. 23. Average precision (AP) for COIL-100 dataset.

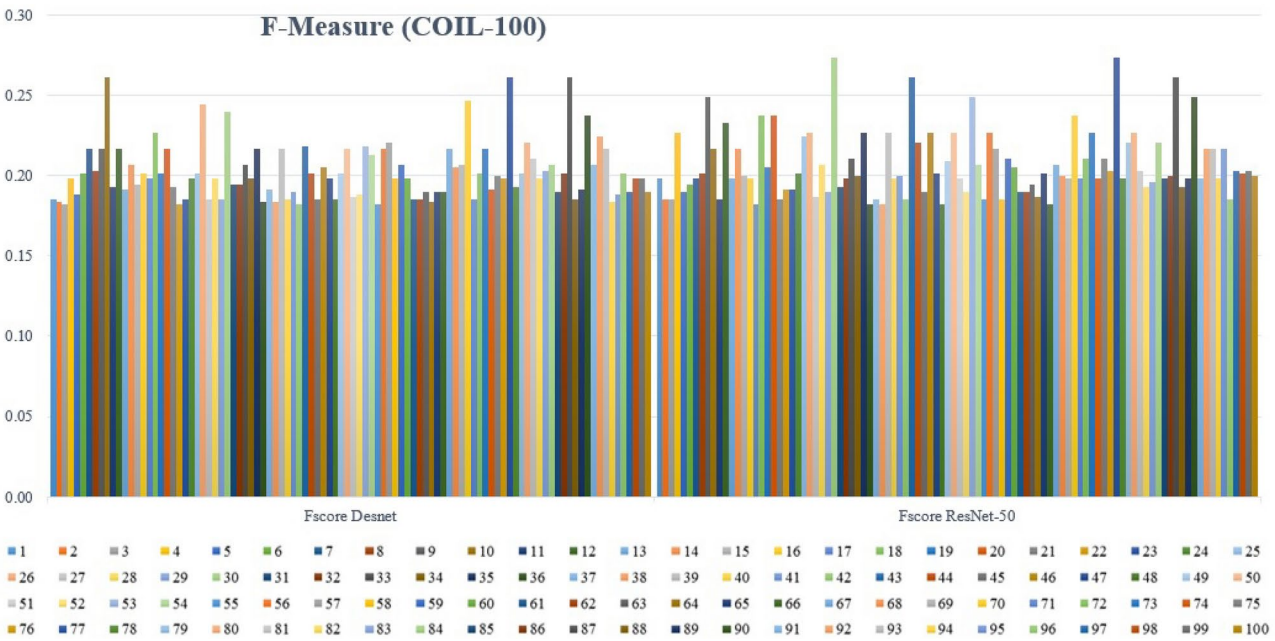


Fig. 24. F-feasure results for COIL-100 dataset.

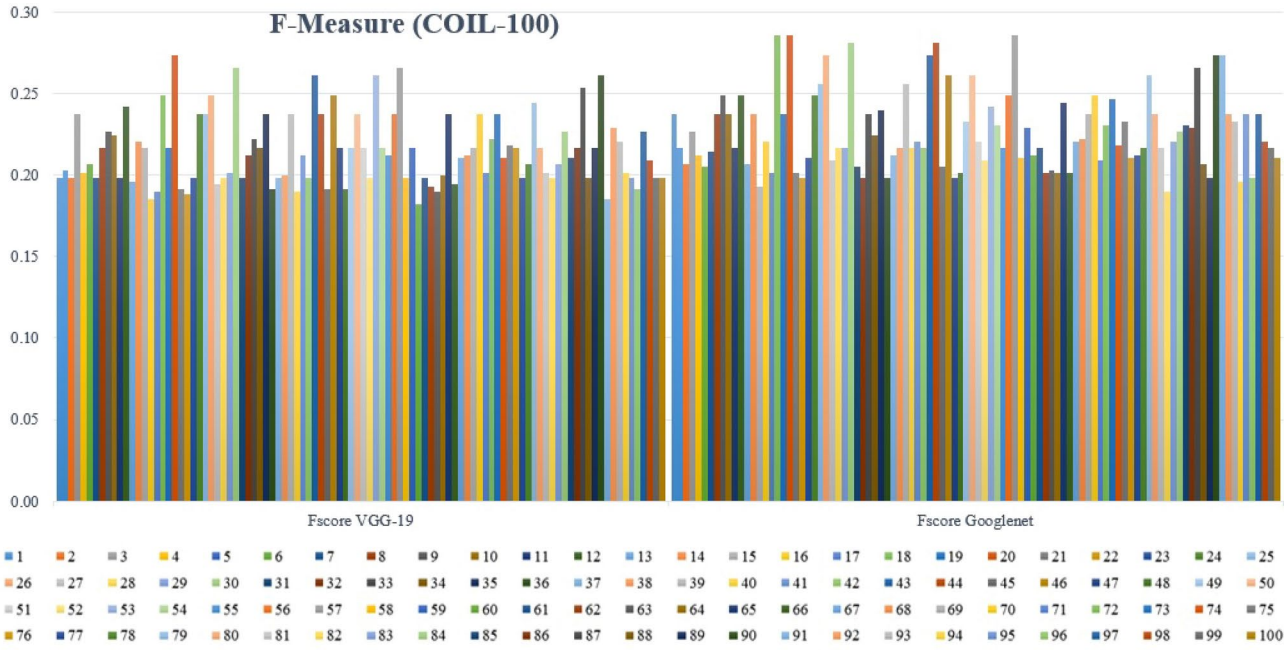


Fig. 25. F-feasure results for COIL-100 dataset.

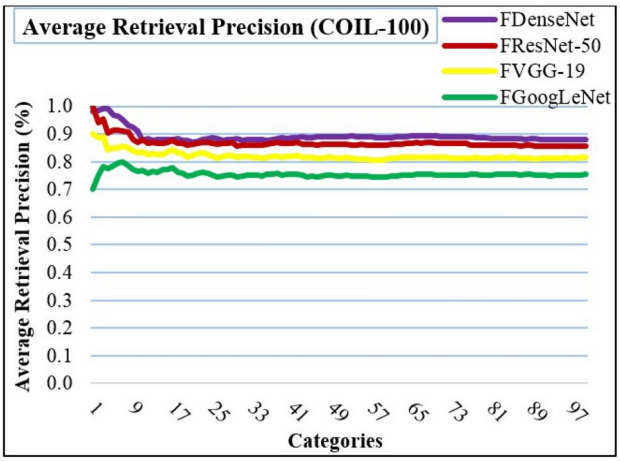


Fig. 26. Average retrieval precision (ARP) for COIL-100 dataset.

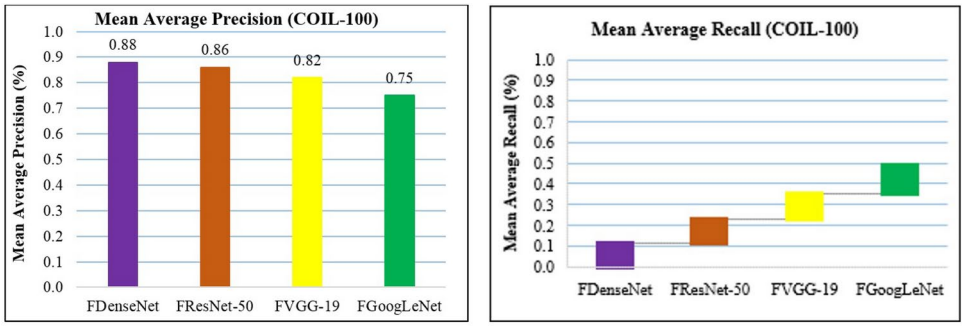
In Fig. 27a, the method illustrates 88% mAP results using FDenseNet, 86% mAP ratios with FResNet-50, 75% for FGoogLeNet, and 82% using FVGG-19 of COIL-100 due to signature influencing, architectural bonding and dataset harmonizing presented algorithms. The presented method shows 0.12 mAR with FResNet-50, 0.11 mAR with FDenseNet, 0.13 with FVGG-19 and 0.14 with FGoogLeNet of COIL-100 dataset in Fig. 27b

Moreover, signature influencing in Fig. 28 is a crucial aspect in the image processing domain which is affectively focused by combining the complex CNN features with architectural presented algorithmic details and color coefficients along with grey level details at deeper descriptor cost. Competitive results are observed in small time fractions for thousands of image comparisons and indexing. Outstanding mean average precision in COIL-100 is focused and perfectly achieved in the signature influencing part of the presented methodology.

Performance on FTVL tropical fruits and 17-FLOWERS datasets

FTVL dataset comprises 2612 vegetables and fruits categories. The FTVL database classes are taiti_lime, spanish_pear, Agata Potato, watermelon, granny_smith_apple, orange, cashew, asterix_potato, diamond_peach, honeydew_melon, plum, fuji_apple, onion, nectarine, and kiwi. Sample images are shown in Fig. 29a.

17-flowers database comprises seventeen image categories; every category includes eighty images. The 17-flowers classes are dandelion, bluebell, iris, tulip, buttercup, cowslip, crocus, tiger lily, windflower, colts foot, daisy, lily valley, sunflower, snowdrop, daffodils, pansy, and fritillary. Sample images are shown in Fig. 29b.



(a) (b)

Fig. 27. **(a)** Mean average precision (mAP) for COIL-100 dataset, and **(b)** Mean average recall (mAR) for COIL-100 dataset.

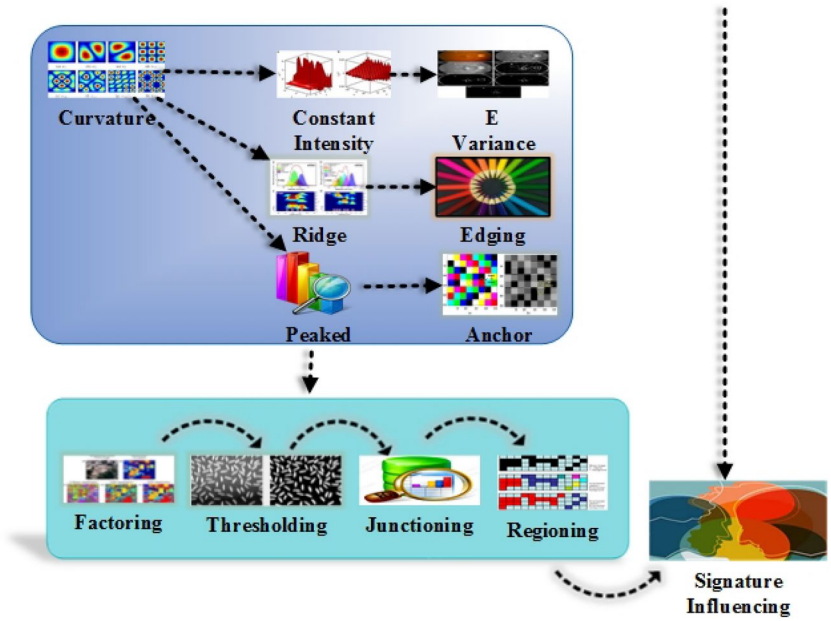


Fig. 28. Signature influencing.

The designed model outperforms average precision, ARP, and f-measure rates for the FTVL dataset in Table 2. It is noticed that the approach provides tremendous results in most of the semantic groups of the FTVL dataset with the same colors and shapes. The color coefficients, distance measuring, thresholding, junctioning, and regioing with convolutional neural network features make it possible to classify images efficiently. The designed approach provides marvelous results using FDenseNet in FTVL categories namely Tahiti Lime, Asteric Potatoes, Kiwi, Diamond Peach, GrannySmith Apple, Astrix, Plum, and Onion. The supremacy of the designed method is to differentiate objects that depend on colors, shapes, and textures which are used to classify images of the FTVL dataset. On the other hand, the presented approach shows the marvelous highest results using FDenseNet in 8/15 tropical fruits semantic groups. The designed method reports outstanding average precision results with FResNet-50 in Cashew, Agata potato, Spanish Pear, Honeydew melon, Nectarine, Watermelon, and Orange image groups of tropical fruits. In Table 2, the presented method also shows significant f-measure ratios of the FTVL database. Therefore, the designed method reports outstanding f-measure ratios with FDenseNet, FResNet-50, FVGG-19, and FGoogLeNet. Furthermore, the presented algorithm reports tremendous ARP results with FDenseNet in most semantic groups which provides improved performance of the presented model of the FTVL dataset. This model presents above 95% ARP ratios by applying features extracted from FDenseNet, above 90% ARP results with FResNet-50, above 85% ARP results with FVGG-19, and 82% ARP results with FGoogLeNet. The presented approach provides 96% mAP with FDenseNet, 95% mAP by FResNet-50, 89% mAP by FVGG-19, and 82% mAP by FGoogLeNet of the FTVL dataset. Architecture bonding can be observed from the experimentation in which the algorithm equally sights 256 semantic groups with high-level mimicking as well as 17 flower groups with high-level color and shape resemblance. The proposed architectural structure of

(a)

(b)

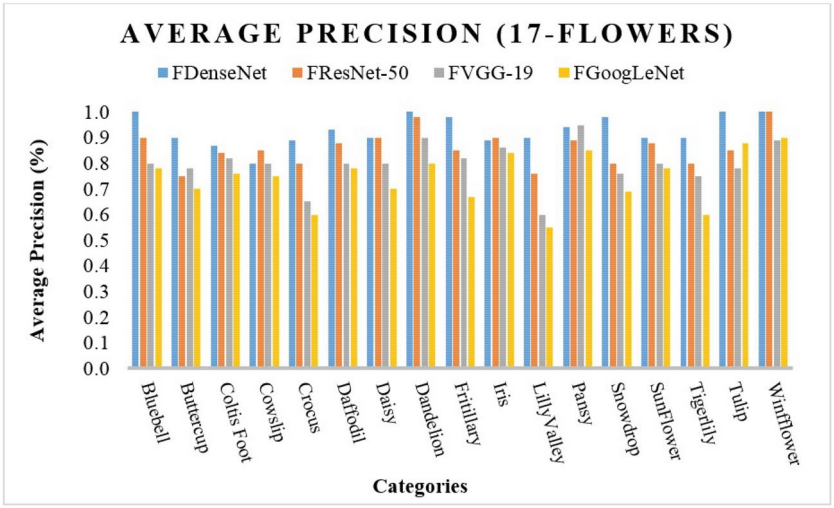
Fig. 29. Sample images from, **(a)** FTVL tropical fruits dataset[5], and **(b)** 17-flowers dataset[57].

	FDenseNet			FResNet			FVGG-19			FGoogleNet		
Category	**AP**	**ARR**	**F**	**AP**	**ARR**	**F**	**AP**	**ARR**	**F**	**AP**	**ARR**	**F**
Taiti Lime	1.00	1.00	0.19	0.95	0.95	0.19	0.90	0.90	0.20	0.80	0.80	0.22
Cashew	0.98	0.99	0.20	0.92	0.94	0.19	0.89	0.89	0.20	0.82	0.81	0.21
Agata Potato	0.88	0.95	0.18	1.00	0.96	0.18	0.95	0.91	0.19	0.90	0.84	0.20
Diamond Peach	0.98	0.96	0.19	0.95	0.96	0.19	0.88	0.91	0.20	0.80	0.83	0.22
GrannySmith Apple	0.98	0.96	0.19	0.94	0.95	0.19	0.91	0.91	0.20	0.85	0.83	0.21
Asterix Potato	1.00	0.97	0.19	0.98	0.96	0.18	0.89	0.90	0.20	0.76	0.82	0.22
Nectarine	0.95	0.97	0.18	0.99	0.96	0.18	0.88	0.90	0.20	0.86	0.83	0.20
Fuji Apple	0.94	0.96	0.20	0.90	0.95	0.20	0.80	0.89	0.22	0.67	0.81	0.24
Watermelon	0.94	0.96	0.19	0.95	0.95	0.19	0.98	0.90	0.18	0.90	0.82	0.20
Honeydew Melon	0.94	0.96	0.18	0.99	0.96	0.18	0.88	0.90	0.20	0.80	0.82	0.22
Spanish Pear	0.98	0.96	0.18	1.00	0.96	0.18	0.94	0.90	0.19	0.84	0.82	0.21
Plum	0.99	0.96	0.20	0.90	0.96	0.20	0.88	0.90	0.20	0.80	0.82	0.22
Kiwi	0.90	0.96	0.20	0.89	0.95	0.20	0.80	0.89	0.22	0.76	0.81	0.22
Onion	0.97	0.96	0.20	0.90	0.95	0.20	0.90	0.89	0.20	0.88	0.82	0.20
Orange	0.98	0.96	0.18	1.00	0.95	0.18	0.93	0.89	0.19	0.89	0.82	0.20

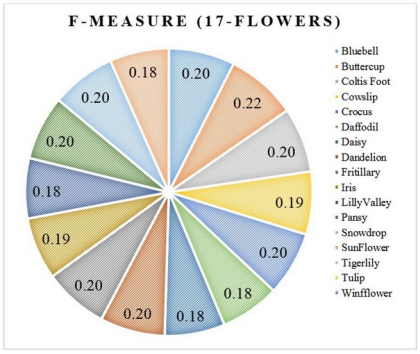
Table 2. AP, ARP, and F-measure rates for FTVL dataset.

CNN is fused with the presented algorithm to focus on the high variety of colors, shapes, and textures similarities at group and feature levels.

In Fig. 30a, the presented method provides marvelous average precision results for the 17-flowers dataset. It is noted that the proposed approach reports tremendous results in most of the semantic groups of 17- flowers with different colors, textures, and shapes. RGB color coefficients and distance measuring with CNN features make it easy to efficient image classification of flowers efficiently. The proposed algorithm shows outstanding AP results using FDenseNet in flower image categories namely Bluebell, Coltis Foot, Snowdrop, Buttercup, Crocus, Dandelion, Fritillary, Daffodil, Lilly Valley, Tiger lily, Sunflower, and Tulip. The robustness and versatility of the model to differentiate objects based on their shapes, colors, and textures shows the important role of classifying flowers. Moreover, the designed approach reports amazingly high results with FDenseNet in the 13/17 flower of 17-flowers dataset. The designed technique provides the highest AP results using FResNet-50 in Cowslip and Iris flower classes. It is noted that the proposed approach reports a high AP ratio with FVGG-19 in the Pansy flower category. In Fig. 30b, the designed approach depicts tremendous f-measure ratios for the 17-flowers dataset. The



(a)



(b)

Fig. 30. **(a)** Average precision for 17-Flowers dataset, and **(b)** F-measure results for COIL-100 dataset for 17-Flowers dataset.

designed method reports f-measures results using FDenseNet, FResNet-50, FVGG-19, and FGoogleNet between 18% and 28% in all semantic groups of the 17-flowers dataset.

The presented algorithm works at color coefficients by applying appropriate displacements at L1 and L2 levels and does not cater to the explanatory aspects of feature sets by only selecting the heads that are thereby connected with the signature influencing dimension. The recent research contribution presents feature fusions which finally provide large feature set sizes that directly influence the experimental cost and there they lose values in merging. However, in the proposed methodology, the signatures are not influenced due to signature length. These are manipulated in such a way that the computational cost is minimal. This research first time introduces diverse CNN complex architectures that have never been used in experimentation before collectively by any researcher. These datasets and CNNs are incorporated with the presented algorithm. This is a heterogeneous architectural bonding that is presented in this research and shows the high level of precision for complex datasets with deep image sensing. In Fig. 31, the color channeling and color-based content retrieval are focused and perfectly achieved in this part of the methodology.

In Fig. 32a, the model provides significant ARP ratios with FDenseNet, FResNet-50, FVGG-19, and FGoogleNet of 17-flowers. In Fig. 32a, the presented approach depicts above 90% ARP results using FDenseNet, above 85% ARP results using FResNet-50, above 79% ARP results by FVGG-19 and above 73% ARP results by FGoogLeNet for most semantic groups of 17-flowers. In Fig. 32b, the proposed approach provides the highest ARP results with FDenseNet. This proposed approach also shows above 0.10 ARR results with FDenseNet, above 0.11 ARR ratios by FResNet-50, above 0.12 ARR results with FVGG-19, and above 0.14 using FGoogLeNet for the 17-flowers dataset.

In Fig. 33a, the presented method illustrates 93% mAP results using FDenseNet, 86% mAP ratios with FResNet-50, 74% for FGoogLeNet, and 80% using FVGG-19 of COIL-100 due to presented algorithms of signature influencing, architectural bonding and dataset harmonizing. The presented method is depicted as 0.12 mAR with FResNet-50, 0.11 mAR with FDenseNet, 0.13 with FVGG-19, and 0.14 with FGoogLeNet of COIL-100 dataset in Fig. 33b. Presented algorithmic channeling is highly emphasized in the 17-flowers dataset which is basically color-based image content analysis.

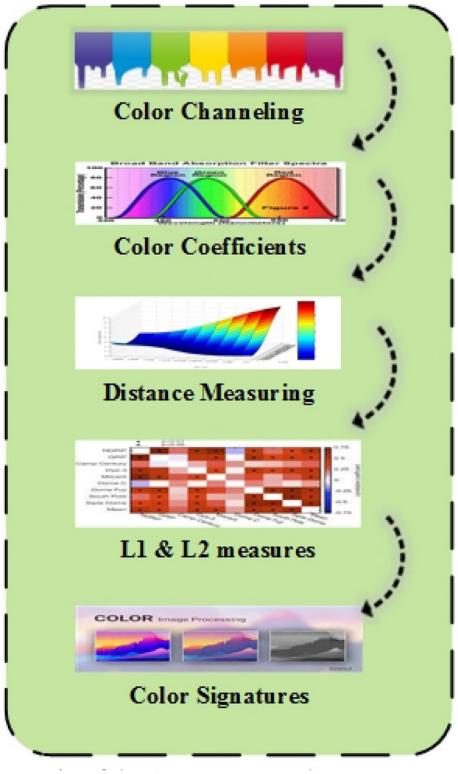


Fig. 31. Signature influencing.

Results of Cifar-10, 17-Flowers, Corel-1000, FTVL, Caltech-256, Corel-10k, Caltech-101 and COIL-100 datasets with predefined research methods

To evaluate the efficiency, robustness, and accuracy of the presented model, it was compared with existing designed approaches. The proposed approach has reported tremendous performance in comparison to existing designed methods. The designed technique presents outstanding average precision results in most semantic groups of the Corel-1000 dataset. The proposed approach provides improved AP ratios in the following categories namely Beaches, Elephants, Africans, Foods, and Mountains. Moreover, mAP results of the presented approach in comparison with state-of-the-art approaches in Fig. 34. The presented method reports 0.96 mAP. DICDLS[59] provides 0.91 mAP at the second highest level. CBRIF[19] reports 0.88 mAP. The ENSR[93] and MSIR[94] report 0.76 mean average precision. DISPR[91], ETRC[95], MDBP[96], MCFG[97], CHOG[98] and FFECR[99] show mean average precision ratios between 0.65 and 0.80. EIRCTS[100] reports the lowest 0.59 mAP. EIRCTS[101] applies the colors and shapes feature so the mAP results are comparatively low when compared with predesigned approaches.

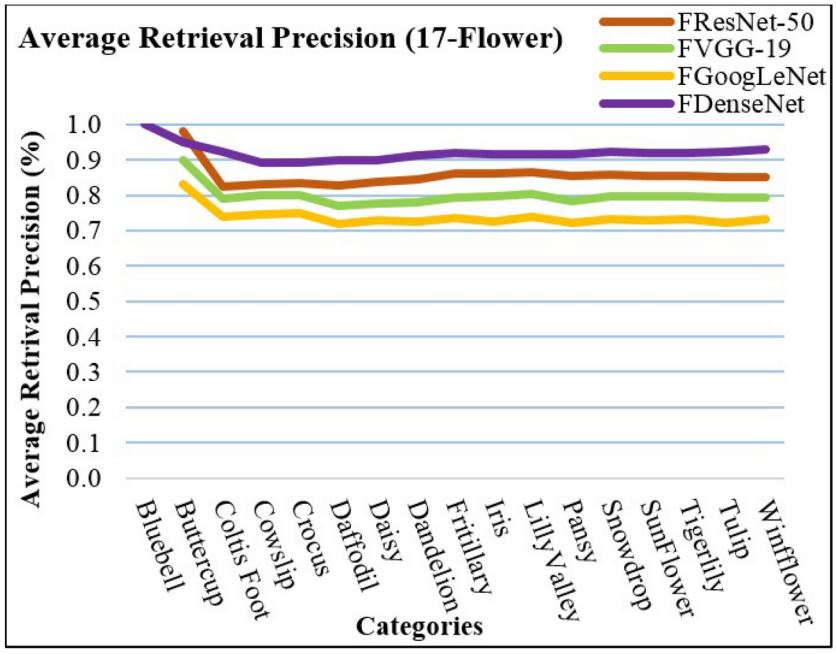
Experimentation is performed for state-of-the-art methods including LBP[102], BRFIF[19], DISBOW[91], HOGHD[103], PTOTM[102], GRITC[104], ORLSIF[105], SRF[106] and RWMSER[107] and results are compared to check the efficiency and accuracy of presented approach. For this, experiments are performed on standard benchmarks including COIL-100, Caltech-256, Corel-10k, and Caltech-101.

The presented approach provides the highest 88% mAP result compared with state-of-the-art approaches for the COIL-100 dataset in Table 3. The mAP results of the presented approach are dominant when compared with state-of-the-art methods. In Fig. 35a, the mAP of the proposed approach is 90% which is 23% higher than DISBOW[91] which covers 67% mAP result at the second level of Caltech-101. Likewise, Caltech-256 image semantic categorization of the presented approach gives 88% highest mAP rates in Fig. 35b. GRITC[104] also covers the textural aspects so the mAP result is comparatively less than other state-of-the-art methods. The proposed approach significantly classifies and indexes the image groups by L2 normalization, distance measuring, and color coefficients.

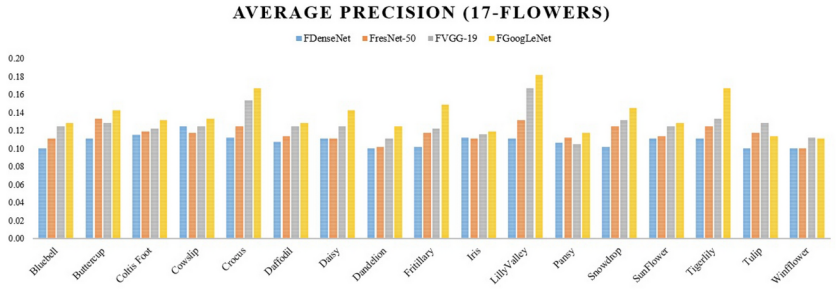
For the Corel-10000 dataset, the presented approach provides outstanding and significantly higher 90% mAP than other existing benchmarks in Table 4. The proposed approach can specifically classify colors and shapes so it reports higher calculated mAP than predefined techniques for the Corel-10k dataset.

The mean average precision results are graphically illustrated in comparison with state-of-the-art techniques of the Cifar-10 dataset for the presented approach and other state-of-the-art methods in Fig. 36. The presented approach outperforms mAP ratios over predefined methods for the Cifar-10 database. The state-of-the-art research approaches namely IETRCI[95], IFFCDIR[108], IPBOMLA[109], IIRSCTS[100], ICBIRCT[110], IMSCBIR[94] and ISPRCNN[111]. The highest mAP results of the presented method as compared with predefined methods are depicted in Fig. 36.

This study also utilized the FTVL dataset due to having images with partial occlusions, illumination differences, pose variations, and cropped objects. Table 5 presents the AP ratios of the FTVL dataset, the designed model

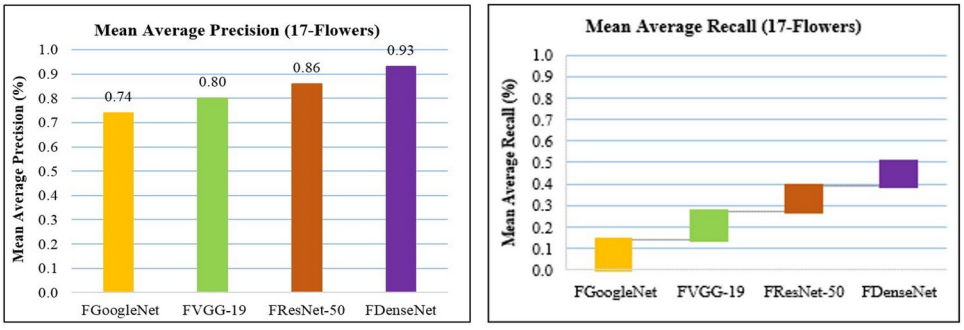


(a)



(b)

Fig. 32. **(a)** Average retrieval precision (ARP) for17-Flowers dataset, and **(b)** Average retrieval recall (ARR) for 17-Flowers dataset.



(a) (b)

Fig. 33. **(a)** Mean average precision (mAP) for the 17-Flowers dataset, and **(b)** Mean average recall (mAR) for the 17-Flowers dataset.

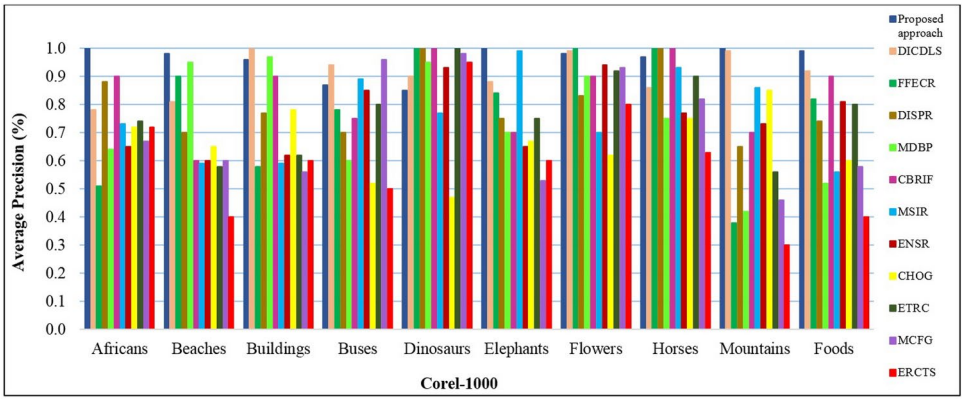
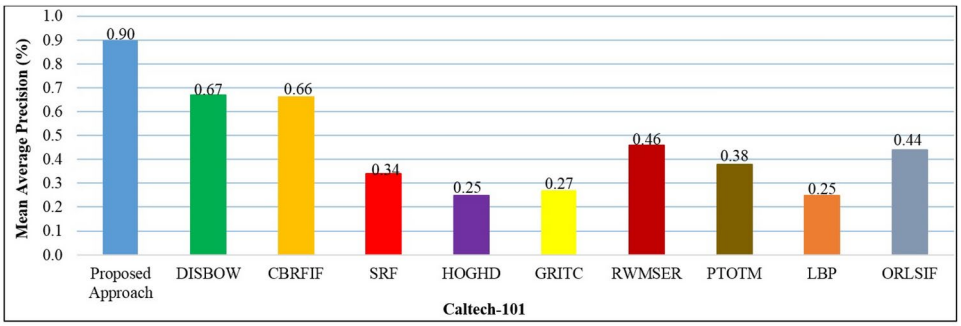


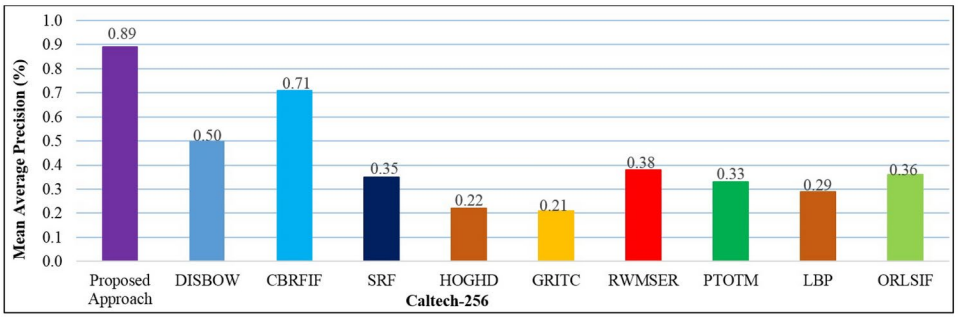
Fig. 34. AP results of designed method versus predefined methods for Corel-1000 dataset.

Proposed	SRF	HOGHD	GRITC	RWMSER	PTOTM	LBP	ORLSIF
0.88	0.49	0.44	0.49	0.52	0.30	0.49	0.26

Table 3. Comparison of designed method map ratios versus predefined methods of COIL-100 dataset.



(a)



(b)

Fig. 35. **(a)** mAP results of designed method versus predefined methods for Caltech-101 dataset, and **(b)** mAP results of designed method versus predefined methods for Caltech-256 dataset.

Proposed	DISBOW	CBRFIF	SRF	HOGHD	GRITC	RWMSER	PTOTM	LBP	ORLSIF
0..92	0.65	0.83	0.26	0.30	0.36	0.57	0.52	0.47	0.25

Table 4. Comparison of designed method mAP ratios versus predefined methods of Corel-10000 dataset.

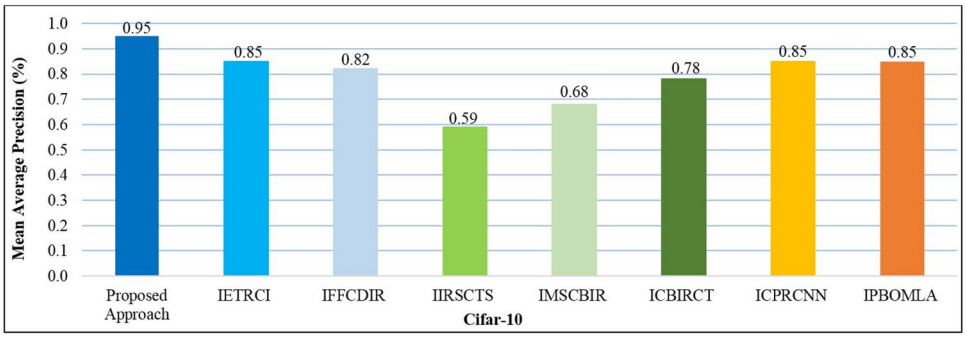


Fig. 36. mAP results of designed model versus predefined methods for Cifar-100 dataset.

Class	Proposed	CDH+SHE	CCV+CLB	CCV+LTP	CBRFF	GCH+LBP	CDH+CLBP	CDH+SHE+CLBP
Agata potato	0.88	0.96	0.99	0.99	0.97	1.00	0.93	0.99
Asterix potato	1.00	0.96	0.98	0.99	1.00	0.96	0.95	0.96
Cashew	0.98	0.99	1.00	1.00	0.99	1.00	1.00	1.00
Diamond peach	0.98	0.89	1.00	1.00	0.91	1.00	0.96	0.96
Fuji apple	0.94	0.60	0.62	0.67	0.93	0.75	0.59	0.71
Granny-Smith apple	0.98	0.96	1.00	1.00	0.97	1.00	0.97	0.99
Honeydew melon	0.94	0.90	0.99	0.99	0.94	0.99	0.99	0.98
Kiwi	0.90	0.75	0.70	0.75	0.95	0.70	0.75	0.77
Nectarine	0.96	0.85	0.84	0.88	0.95	0.87	0.81	0.90
Onion	0.97	0.99	1.00	1.00	0.89	1.00	1.00	1.00
orange	0.98	0.98	0.98	0.99	0.96	0.99	0.98	0.98
Plum	0.99	0.98	0.98	1.00	0.90	0.99	0.97	0.99
Spanish pear	0.98	0.70	0.85	0.75	0.91	0.84	0.83	0.86
Taiti lime	1.00	0.96	1.00	1.00	1.00	1.00	1.00	0.99
Watermelon	0.94	0.99	0.99	1.00	0.94	0.99	0.98	0.99

Table 5. Comparison of AP of designed model average precision ratios versus predefined methods of FTVL dataset.

Approach	Average precision
CDH+SHE	0.90
CCV+CLB	0.93
CCV+LTP	0.93
CBRFF	0.95
GCH+LBP	0.94
CDH+CLBP	0.91
CDH+SHE+CLBP	0.94
Proposed	0.96

Table 6. Performance comparison for average precision using the FTVL dataset.

is compared with predefined methods in CBRFF[5] and[112]. The presented research approach outperforms AP results in most of the semantic categories of the FTVL database. Few predefined approaches CDH +SEH show less accurateness because of its missing nature in their approaches to crop the objects. Some methods collect the textural attributes and report the average precision. Therefore, objects with similar colors and shapes are quite hard to recognize images. The proposed model uses color coordinates by color, shape, and texture properties to give 0.96 mean average precision.

To provide a better picture of the results from the proposed approach in comparison to existing research works, Table 6 provides averaged results for the FTVL dataset. The dataset contains 15 classes and class-wide results are given in Table 5. Averaged results for the FTVL dataset indicate that the proposed approach provides robust results with enhanced average precision, compared to existing methods in the FTVL dataset.

Discussion

A new framework for image feature extraction builds upon multi-channel image processing through state-of-the-art architectural approaches including DenseNet ResNet-50 VGG-19 and GoogLeNet. The proposed method works across color and grayscale formats through displaced color channels combined with spatial heads so features maintain deep and complex representations. The new method provides the model with the ability to process diverse complex datasets which leads to stronger accuracy than traditional single-channel image processing approaches. The primary achievement within this research is combining local-global vectors and regioned features to achieve complete image data representation. The proposed method merges local and global features into one consolidated feature extraction system that maintains an overall balanced feature extraction practice. The combination of local and global processing techniques enhances the accuracy of identification tasks when applied to Caltech-101, Cifar-10, and 17-Flowers databases. The proposed approach used observation auto-correlation to handle anisotropic noise while extracting correlated shape parameters which output responses with peaked curvatures and ridges as well as peaks. Factor threshold and region processes are applied to these features to produce more reliable image representations. The approach delivers exceptional results when handling data collections containing intricate visual features since traditional CNNs fail to abstract significant characteristics from complex or noisy inputs.

This study demonstrates through extensive examination that the proposed approach delivers better performance than existing systems on a range of benchmark datasets. A substantial enhancement in retrieval accuracy became apparent when comparing the approach to conventional CNN models on Cifar-10. The enhanced performance results from synergizing multi-channel image processing with multiple CNN architectures that deliver optimized features along with effective learning dynamics. Filtering images with color information throughout retrieval leads to better performance, especially on object datasets that need color discrimination such as FTVL Tropical Fruits and Zubud. The proposed method achieves better accuracy and precision together with higher recall performance than three prominent architecture frameworks including ResNet-50, VGG-19, and GoogLeNet. By combining several CNN models with the proposed channeled algorithm, we have obtained better and more adaptable image feature representations than traditional networks using single architectural frameworks. In this way, the proposed technique stands out in multi-class settings and image retrieval applications due to its ability to integrate diverse high-dimensional features from different images effectively. This method solves current limitations that exist within semantic feature extraction methods. Traditional datasets must rely on conventional single-channel representations and conventional CNNs because they do not work effectively with complex multiple-dimensional datasets. The current work implements multi-architecture integration and advanced color channel processing to achieve better image feature extraction with capabilities extending across multiple image sensing operations. The profound impact arises for image retrieval, object recognition, and semantic segmentation applications since these require high-precision efficient feature representations.

The combination of multiple CNN architectures with innovative feature fusion techniques introduced in this work strengthens deep learning's growing ability to analyze images. The incorporation of color and grayscale features together with displaced color techniques and spatial heads differentiates the proposed approach from other methods leading to its unique position for image analysis work in the future. The results of this research advance the field and create foundational capabilities for additional studies examining hybrid CNN architectures linked with multi-channel image processing to service various image-based applications.

Conclusion

This research presents an extraordinary fused convolutional neural network approach that joins the power of GoogLeNet, DenseNet, VGG-19, and ResNet-50 with the presented algorithmic image retrieval model. This versatile classification and retrieval methodology is capable of detecting overlayed, cluttered, background, and foreground image contents with content synthesis and deep analysis. This research is aimed at compact and efficient image feature vector extraction to deeply learn the images parallel at colored and grey integrations along with CNN signatures, with channeled algorithm integrations to have uniformity at diverse datasets and semantic groups. Novel signature influencing is obtained by the fusion of bounded, primitive-parametrized, local-global, and anchored vectors with regioned features to ultimately generate harmonized signatures which are thereby indexed with KNN. The competitive results endorsed the superiority of the proposed method by its comparison with standard benchmarks of Caltech-101, Cifar-10, Caltech-256, Cifar-100, Corel-10000, 17-Flowers, COIL-100, FTVL Tropical Fruits, Corel-1000, and Zubud. This research work can be extended to DesNet, MobileNet, and Deep cloud-based architecture.

Data availability

The datasets are publicly available at the following links: Cifar-10 https://www.kaggle.com/c/cifar-10. Corel-10k https://www.kaggle.com/datasets/michelwilson/corel10k. COIL-100 https://www.kaggle.com/datasets/jessicali9530/coil100.

Received: 15 December 2024; Accepted: 14 February 2025
Published online: 04 March 2025

References

1. Ajani, S. N. et al. Advancements in computing: Emerging trends in computational science with next-generation computing. *International Journal of Intelligent Systems and Applications in Engineering* **12**(7s), 546–559 (2024).

2. Wang, Y., Wang, S. & Tan, M. Path generation of autonomous approach to a moving ship for unmanned vehicles. *IEEE Transactions on Industrial Electronics* **62**(9), 5619–5629 (2015).
3. Scaramuzza, D. et al. Vision-controlled micro flying robots: from system design to autonomous navigation and mapping in gps-denied environments. *IEEE Robotics & Automation Magazine* **21**(3), 26–40 (2014).
4. Banerjee, A. Comparative analysis on hybrid content & context-basedimage retrieval system. *Journal of Information Technology Management 13(Special Issue: Role of ICT in Advancing Business and Management)*, 133–142 (2021).
5. Nayagam, M. G. & Ramar, K. Reliable object recognition system for cloud video data based on ldp features. *Computer Communications* **149**, 343–349 (2020).
6. Douze, M., Revaud, J., Verbeek, J., Jégou, H. & Schmid, C. Circulant temporal encoding for video retrieval and temporal alignment. *International Journal of Computer Vision* **119**, 291–306 (2016).
7. Ranjan, R., Patel, V. M., Chellappa, R. & Castillo, C. D. Deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *Google Patents*. US Patent 10,860,837 (2020).
8. Dong, Y., Tao, D., Li, X., Ma, J. & Pu, J. Texture classification and retrieval using shearlets and linear regression. *IEEE transactions on cybernetics* **45**(3), 358–369 (2014).
9. Bergamo, A. & Torresani, L. Classemes and other classifier-based features for efficient object categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **36**(10), 1988–2001 (2014).
10. Bhaskaranand, M. & Gibson, J. D. Global motion assisted low complexity video encoding for uav applications. *IEEE Journal of Selected Topics in Signal Processing* **9**(1), 139–150 (2014).
11. Sadique, M. F., Biswas, B. K. & Haque, S. R. Unsupervised content-based image retrieval technique using global and local features. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pp. 1–6 (2019). IEEE
12. Qing, Q. Aggregating local features of convolutional neural network for material image retrieval. In *Journal of Physics: Conference Series*, vol. 1948, p. 012061 (2021). IOP Publishing.
13. Zope, S. & Patil, D. D. A review on cbir analyzing various feature extraction techniques and distance metrics (2020).
14. Sipiran, I. & Bustos, B. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer* **27**, 963–976 (2011).
15. Uliyan, D. M., Jalab, H. A., & Wahab, A. W. A. Copy move image forgery detection using hessian and center symmetric local binary pattern. In *2015 IEEE Conference on Open Systems (ICOS)*, pp. 7–11 (2015). IEEE
16. Kasiselvanathan, M., Sangeetha, V. & Kalaiselvi, A. Palm pattern recognition using scale invariant feature transform. *International Journal of Intelligence and Sustainable Computing* **1**(1), 44–52 (2020).
17. Mikolajczyk, K. & Schmid, C. Scale & affine invariant interest point detectors. *International journal of computer vision* **60**, 63–86 (2004).
18. Shi, X. & Qian, X. Exploring spatial and channel contribution for object based image retrieval. *Knowledge-Based Systems* **186**, 104955 (2019).
19. Ahmed, K. T., Ummesafi, S. & Iqbal, A. Content based image retrieval using image features information fusion. *Information Fusion* **51**, 76–99 (2019).
20. Guo, J.-M., Prasetyo, H. & Chen, J.-H. Content-based image retrieval using error diffusion block truncation coding features. *IEEE Transactions on Circuits and Systems for Video Technology* **25**(3), 466–481 (2014).
21. Singh, J., Bajaj, A., Mittal, A., Khanna, A. & Karwayun, R. Content based image retrieval using gabor filters and color coherence vector. In *2018 IEEE 8th International Advance Computing Conference (IACC)*, pp. 290–295 (2018). IEEE
22. Saritha, R. R., Paul, V. & Kumar, P. G. Content based image retrieval using deep learning process. *Cluster Computing* **22**, 4187–4200 (2019).
23. Chen, C., Zhang, B., Su, H., Li, W. & Wang, L. Land-use scene classification using multi-scale completed local binary patterns. *Signal, image and video processing* **10**, 745–752 (2016).
24. Krizhevsky, A., Sutskever, I., & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, **25** (2012)
25. Anubha Pearline, S., Sathiesh Kumar, V. & Harini, S. A study on plant recognition using conventional image processing and deep learning approaches. *Journal of Intelligent & Fuzzy Systems* **36**(3), 1997–2004 (2019).
26. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
27. Zhu, Y. & Newsam, S. Densenet for dense flow. In *2017 IEEE International Conference on Image Processing (ICIP)*, pp. 790–794 (2017). IEEE
28. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9 (2015).
29. Imbriaco, R., Sebastian, C., Bondarev, E. & With, P. H. Aggregated deep local features for remote sensing image retrieval. *Remote Sensing* **11**(5), 493 (2019).
30. Lopes, N. & Ribeiro, B. Towards adaptive learning with improved convergence of deep belief networks on graphics processing units. *Pattern recognition* **47**(1), 114–127 (2014).
31. Johnson, R. & Zhang, T. Semi-supervised convolutional neural networks for text categorization via region embedding. *Advances in neural information processing systems*, **28** (2015)
32. Aslam, N., Rustam, F., Lee, E., Washington, P. B. & Ashraf, I. Sentiment analysis and emotion detection on cryptocurrency related tweets using ensemble lstm-gru model. *Ieee Access* **10**, 39313–39324 (2022).
33. Abdel-Hamid, O. et al. Convolutional neural networks for speech recognition. *IEEE/ACM Transactions on audio, speech, and language processing* **22**(10), 1533–1545 (2014).
34. Borji, A., Cheng, M.-M., Jiang, H. & Li, J. Salient object detection: A benchmark. *IEEE transactions on image processing* **24**(12), 5706–5722 (2015).
35. Zhou, W., Li, H. & Tian, Q. Recent advance in content-based image retrieval: A literature survey. arXiv preprint arXiv:1706.06064 (2017)
36. Kim, Y. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
37. Abdeljaber, O. et al. 1-d cnns for structural damage detection: Verification on a structural health monitoring benchmark data. *Neurocomputing* **275**, 1308–1317 (2018).
38. Kanimozhi, T. & Latha, K. An integrated approach to region based image retrieval using firefly algorithm and support vector machine. *Neurocomputing* **151**, 1099–1111 (2015).
39. Kavitha, H. & Sudhamani, M. Image retrieval based on object recognition using the harris corner and edge detection technique. In *International Conference on Communication, VLSI & Signal Processing*, pp. 181–184 (2013).
40. Kavitha, H. & Sudhamani, M. Content-based image retrieval using edge and gradient orientation features of an object in an image from database. *Journal of Intelligent Systems* **25**(3), 441–454 (2016).
41. Kaur, S. & Verma, M. P. Content based image retrieval: integration of neural networks using speed-up robust feature and svm. *International Journal of Computer Science and Information Technologies* **6**(1), 243–248 (2015).
42. Shanmugam, B., Rathinavel, R., Perumal, T. & Subbaiyan, S. An efficient perceptual of cbir system using mil-svm classification and surf feature extraction. *The International Arab Journal of Information Technology* **14**(4), 428–435 (2017).

43. Hassan, M. U., Shohag, M. S. A., Niu, D., Shaukat, K., Zhang, M., Zhao, W. & Zhao, X. A framework for the revision of large-scale image retrieval benchmarks. In *Eleventh International Conference on Digital Image Processing (ICDIP 2019)*, vol. 11179, pp. 1154–1161 (2019). SPIE
44. Putzu, L., Piras, L. & Giacinto, G. Convolutional neural networks for relevance feedback in content based image retrieval: A content based image retrieval system that exploits convolutional neural networks both for feature extraction and for relevance feedback. *Multimedia Tools and Applications* **79**, 26995–27021 (2020).
45. Lyra, L. O., Fabris, A. E. & Florindo, J. B. A multilevel pooling scheme in convolutional neural networks for texture image recognition. *Applied Soft Computing*, 111282 (2024)
46. Hu, R. & Collomosse, J. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *Computer Vision and Image Understanding* **117**(7), 790–806 (2013).
47. Liu, Y., Chen, W., Qu, H., Mahmud, S. H. & Miao, K. Spatial division networks for weakly supervised detection. *Neural Computing and Applications* **33**, 4965–4978 (2021).
48. Alluri, L. & Dendukuri, H. An efficient system for cbir using deep learning convolutional neural networks. *Int J Recent Dev Sci Technol* **4**(1), 160–167 (2020).
49. Sezavar, A., Farsi, H. & Mohamadzadeh, S. Content-based image retrieval by combining convolutional neural networks and sparse representation. *Multimedia Tools and Applications* **78**, 20895–20912 (2019).
50. Byambajav, B., Alikhanov, J., Fang, Y., Ko, S. & Jo, G. S. Transfer learning using multiple convnet layers activation features with principal component analysis for image classification. *J Intell Inform Syst* **24**(1), 205–225 (2018).
51. Chan, T.-H. et al. Pcanet: A simple deep learning baseline for image classification?. *IEEE transactions on image processing* **24**(12), 5017–5032 (2015).
52. Xia, R., Pan, Y., Lai, H., Liu, C., & Yan, S. Supervised hashing for image retrieval via image representation learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 28 (2014)
53. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**(1), 1929–1958 (2014).
54. Kuo, C.-H., Chou, Y.-H. & Chang, P.-C. Using deep convolutional neural networks for image retrieval. *Electronic Imaging* **2016**(2), 1–6 (2016).
55. Kanwal, K., Ahmad, K. T., Khan, R., Abbasi, A. T. & Li, J. Deep learning using symmetry, fast scores, shape-based filtering and spatial mapping integrated with cnn for large scale image retrieval. *Symmetry* **12**(4), 612 (2020).
56. Lavinia, Y., Vo, H. H., & Verma, A. Fusion based deep cnn for improved large-scale image action recognition. In *2016 IEEE International Symposium on Multimedia (ISM)*, pp. 609–614 (2016). IEEE.
57. Yamamoto, S., Nishimura, T., Akagi, Y., Takimoto, Y., Inoue, T. & Toda, H. Pbg at the ntcir-13 lifelog-2 lat, lsat, and lest tasks. In *NTCIR* (2017)
58. Mateen, M., Wen, J., Nasrullah, Song, S. & Huang, Z. Fundus image classification using vgg-19 architecture with pca and svd. *Symmetry* 11(1), 1 (2018)
59. Kanwal, K., Tehseen Ahmad, K., Khan, R., Alhusaini, N. & Jing, L. Deep learning using isotroping, laplacing, eigenvalues interpolative binding, and convolved determinants with normed mapping for large-scale image retrieval. *Sensors* **21**(4), 1139 (2021).
60. Zhou, M. et al. Perception-oriented u-shaped transformer network for 360-degree no-reference image quality assessment. *IEEE Transactions on Broadcasting* **69**(2), 396–405 (2023).
61. Ma, D. et al. Transformer-optimized generation, detection, and tracking network for images with drainage pipeline defects. *Computer-Aided Civil and Infrastructure Engineering* **38**(15), 2109–2127 (2023).
62. Wang, W. et al. Low-light image enhancement based on virtual exposure. *Signal Processing: Image Communication* **118**, 117016 (2023).
63. Zhou, G., Qian, L. & Gamba, P. A novel iterative self-organizing pixel matrix entanglement classifier for remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing* (2024)
64. Zhou, Z., Li, Z., Zhou, W., Chi, N., Zhang, J. & Dai, Q. Resource-saving and high-robustness image sensing based on binary optical computing. *Laser & Photonics Reviews*, 2400936 (2024)
65. Zhou, G. & Liu, X. Orthorectification model for extra-length linear array imagery. *IEEE Transactions on Geoscience and Remote Sensing* **60**, 1–10 (2022).
66. Zhou, G. et al. Orthorectification of fisheye image under equidistant projection model. *Remote Sensing* **14**(17), 4175 (2022).
67. Wang, Z., Zhang, Z., Qi, W., Yang, F. & Xu, J. Freqgan: Infrared and visible image fusion via unified frequency adversarial learning. *IEEE Transactions on Circuits and Systems for Video Technology* (2024)
68. Shi, G., Deng, S., Wang, B., Feng, C., Zhuang, Y., & Wang, X. One for all: A unified generative framework for image emotion classification. *IEEE Transactions on Circuits and Systems for Video Technology* (2023)
69. Luo, H., Zhang, Q., Sun, G., Yu, H., & Niyato, D. Symbiotic blockchain consensus: Cognitive backscatter communications-enabled wireless blockchain consensus. *IEEE/ACM Transactions on Networking* (2024)
70. Dai, M., Sun, G., Yu, H., Wang, S. & Niyato, D. User association and channel allocation in 5g mobile asymmetric multi-band heterogeneous networks. arXiv preprint arXiv:2405.18797 (2024)
71. Yang, X. et al. ulidr: An inertial-assisted unmodulated visible light positioning system for smartphone-based pedestrian navigation. *Information Fusion* **113**, 102579 (2025).
72. Wang, E., Yang, Y., Wu, J., Liu, W. & Wang, X. An efficient prediction-based user recruitment for mobile crowdsensing. *IEEE Transactions on Mobile Computing* **17**(1), 16–28 (2017).
73. Ma, J. et al. Robust feature matching for remote sensing image registration via locally linear transforming. *IEEE Transactions on Geoscience and Remote Sensing* **53**(12), 6469–6481 (2015).
74. Karczmarek, P., Kiersztyn, A., Pedrycz, W. & Dolecki, M. An application of chain code-based local descriptor and its extension to face recognition. *Pattern Recognition* **65**, 26–34 (2017).
75. Shakhnoza, M., Sabina, U., Sevara, M. & Cho, Y.-I. Novel video surveillance-based fire and smoke classification using attentional feature map in capsule networks. *Sensors* **22**(1), 98 (2021).
76. Yue, S., Zeng, S., Liu, L., Eldar, Y.C. & Di, B. Hybrid near-far field channel estimation for holographic mimo communications. *IEEE Transactions on Wireless Communications* (2024)
77. Liu, Z., Xiong, X., Li, Y., Yu, Y., Lu, J., Zhang, S. & Xiong, F. Hygloadattack: Hard-label black-box textual adversarial attacks via hybrid optimization. *Neural Networks*, 106461 (2024)
78. Fan, H., Wang, C. & Li, S. Novel method for reliability optimization design based on rough set theory and hybrid surrogate model. *Computer Methods in Applied Mechanics and Engineering* **429**, 117170 (2024).
79. Chu, H., Pan, X., Jiang, J., Li, X. & Zheng, L. Adaptive and robust channel estimation for irs-aided millimeter-wave communications. *IEEE Transactions on Vehicular Technology* (2024)
80. Shi, M. et al. Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. *Mechanical Systems and Signal Processing* **188**, 110022 (2023).
81. Zhao, X. et al. Target-driven visual navigation by using causal intervention. *IEEE Transactions on Intelligent Vehicles* **9**(1), 1294–1304 (2023).
82. Wang, T. et al. Reslnet: deep residual lstm network with longer input for action recognition. *Frontiers of Computer Science* **16**(6), 166334 (2022).

83. Chu, L. et al. A deniable encryption method for modulation-based dna storage. *Interdisciplinary Sciences: Computational Life Sciences* **16**(4), 872–881 (2024).
84. Li, X., Lu, Z., Yuan, M., Liu, W., Wang, F., Yu, Y. & Liu, P. Tradeoff of code estimation error rate and terminal gain in scer attack. *IEEE Transactions on Instrumentation and Measurement* (2024)
85. Guo, T., Yuan, H., Hamzaoui, R., Wang, X. & Wang, L. Dependence-based coarse-to-fine approach for reducing distortion accumulation in g-pcc attribute compression. *IEEE Transactions on Industrial Informatics* (2024)
86. Jing, L., Fan, X., Feng, D., Lu, C. & Jiang, S. A patent text-based product conceptual design decision-making approach considering the fusion of incomplete evaluation semantic and scheme beliefs. *Applied Soft Computing* **157**, 111492 (2024).
87. Huang, C., Wang, Y., Jiang, Y., Li, M., Huang, X., Wang, S., Pan, S., & Zhou, C. Flow2gnn: Flexible two-way flow message passing for enhancing gnns beyond homophily. *IEEE Transactions on Cybernetics* (2024)
88. Li, T., Long, Q., Chai, H., Zhang, S., Jiang, F., Liu, H., Huang, W., Jin, D., & Li, Y. Generative ai empowered network digital twins: Architecture, technologies, and applications. *ACM Computing Surveys* (2025)
89. Jiang, F., Li, T., Lv, X., Rui, H. & Jin, D. Physics-informed neural networks for path loss estimation by solving electromagnetic integral equations. *IEEE Transactions on Wireless Communications* (2024)
90. Xu, S., Chou, W. & Dong, H. A robust indoor localization system integrating visual localization aided by cnn-based image retrieval with monte carlo localization. *Sensors* **19**(2), 249 (2019).
91. Ahmed, K. T., Afzal, H., Mufti, M. R., Mehmood, A. & Choi, G. S. 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* **8**, 90351–90379 (2020).
92. Nene, S. A., Nayar, S. K. & Murase, H. et al. Columbia object image library (coil-20) (1996)
93. Irtaza, A., Jaffar, M. A., Aleisa, E. & Choi, T.-S. Embedding neural networks for semantic association in content based image retrieval. *Multimedia tools and applications* **72**, 1911–1931 (2014).
94. ElAlami, M. E. A new matching strategy for content based image retrieval system. *Applied Soft Computing* **14**, 407–418 (2014).
95. Shrivastava, N. & Tyagi, V. An efficient technique for retrieval of color images in large databases. *Computers & Electrical Engineering* **46**, 314–327 (2015).
96. Dubey, S. R., Singh, S. K. & Singh, R. K. Multichannel decoded local binary patterns for content-based image retrieval. *IEEE transactions on image processing* **25**(9), 4018–4032 (2016).
97. Xiao, Y., Wu, J. & Yuan, J. mcentrist: A multi-channel feature generation mechanism for scene categorization. *IEEE Transactions on Image Processing* **23**(2), 823–836 (2013).
98. Pan, S., Sun, S., Yang, L., Duan, F. & Guan, A. Content retrieval algorithm based on improved hog. In *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*, pp. 438–441 (2015). IEEE
99. Walia, E. & Pal, A. Fusion framework for effective color image retrieval. *Journal of Visual Communication and Image Representation* **25**(6), 1335–1348 (2014).
100. Wang, X.-Y., Yu, Y.-J. & Yang, H.-Y. An effective image retrieval scheme using color, texture and shape features. *Computer Standards & Interfaces* **33**(1), 59–68 (2011).
101. Arora, S., Bhaskara, A., Ge, R. & Ma, T. Provable bounds for learning some deep representations. In *International Conference on Machine Learning*, pp. 584–592 (2014). PMLR
102. Ojala, T., Pietikainen, M., & Harwood, D.: Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Proceedings of 12th International Conference on Pattern Recognition*, vol. 1, pp. 582–585 (1994). IEEE
103. Dalal, N. & Triggs, B. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893 (2005). Ieee
104. Ojala, T., Pietikainen, M. & Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence* **24**(7), 971–987 (2002).
105. Lowe, D. G. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150–1157 (1999). Ieee
106. Bay, H., Tuytelaars, T. & Van Gool, L. Surf: Speeded up robust features. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006*. Proceedings, Part I 9, pp. 404–417 (2006). Springer
107. Matas, J., Chum, O., Urban, M. & Pajdla, T. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing* **22**(10), 761–767 (2004).
108. Raghuwanshi, G. & Tyagi, V. Feed-forward content based image retrieval using adaptive tetrolet transforms. *Multimedia Tools and Applications* **77**(18), 23389–23410 (2018).
109. Snoek, J., Larochelle, H., Adams, R.P.: Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems*, **25** (2012)
110. Lin, C.-H., Chen, R.-T. & Chan, Y.-K. A smart content-based image retrieval system based on color and texture feature. *Image and vision Computing* **27**(6), 658–665 (2009).
111. Zeiler, M. D. & Fergus, R.: *Stochastic pooling for regularization of deep convolutional neural networks*. arXiv preprint arXiv:1301.3557 (2013)
112. Dubey, S. R. & Jalal, A. S. Fruit and vegetable recognition by fusing colour and texture features of the image using machine learning. *International journal of applied pattern recognition* **2**(2), 160–181 (2015).

Author contributions

KK conceptualization, formal analysis, writing - the original manuscript. KTA conceptualization, data curation, writing - the original manuscript. AS methodology, data curation, formal analysis. LJ supervision, methodology, project administration. HG funding acquisition, investigation, visualization. LEPG software, project administration, resources. HK visualization, investigation, software IA supervision, validation, writing - review & edit the manuscript. All authors reviewed the manuscript and approved it.

Funding

This research was supported by the European University of the Atlantic. This study was also funded by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2025R192), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to K.K., H.K. or I.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.