Received: October 26, 2024. Revised: November 14, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Automatic Image Annotation and Retrieval Using Fuzzy C-Means Clustering with Gaussian Bare Bones White Shark Optimization

Janani Shivashankar[1]* Jahnavi Shivashankar[2] Anil Kumar Bhaskar[3]
Manjunath Doddegowdanakppalu Rajegowda[4] Sampath Satheesh Kumar[5] Shiva Sumanth Reddy[6]

[1]*Department of Computer Science and Engineering, K.H. Kabbur Institute of Engineering, Dharwad, India*
[2]*Department of Artificial Intelligence and Machine Learning, B.M.S College of Engineering, Bengaluru, India*
[3]*Department of Computer Science and Engineering, RV Institute of Technology and Management, Bengaluru, India*
[4]*Department of Computer Science and Engineering (IoT & Cybersecurity),*
B.M.S College of Engineering, Bengaluru, India
[5]*Department of Computer Science, REVA University, Bengaluru, India*
[6]*Department of Computer Science and Engineering,*
Dayananda Sagar Academy of Technology and Management, Bengaluru, India
* Corresponding author's Email: janani.cse.jss@gmail.com

Abstract: Image annotation is the process of assigning significant labels to specific parts of an image, making it easier for systems to categorize, interpret, and retrieve visual data. These annotations are then used in the retrieval process to search for and recover images based on tagged data, improving the accuracy and efficiency of identifying relevant images. However, inadequate or inconsistent annotations lead to poor image retrieval performance, decreasing the overall effectiveness. This research proposes the Fuzzy C-Means Clustering-Gaussian Bare Bones White Shark Optimization (FCM-GBWSO) algorithm for automatic image annotation and retrieval. In WSO, GB is incorporated to prevent premature convergence and avoid local optima issues in generating cluster centroids and clustered images. Various feature extraction techniques, such as ResNet50 and Color Moments, are applied to extract features effectively. Feature transformation and Neighbourhood Component Analysis (NCA) are then used to transform the features into similar significance and select the best features. The proposed FCM-GBWSO achieves an average precision of 0.98, 0.96, and 0.97 on the Corel-10k, Caltech 256 datasets, and Corel 1k respectively, outperforming existing methods like hierarchical clustering and Deep Neural Network-based Deep Search and Rescue (DNN-SAR).

Keywords: Color moments, Fuzzy c-means clustering-gaussian bare bones white shark optimization, Neighbourhood component analysis, Retrieval, ResNet50.

1. Introduction

Image annotation is the process of labeling or tagging images with descriptive information to recognize and categorize objects within the image. Image retrieval is the process of searching for and identifying appropriate images from vast datasets based on input queries. Searching for an image based on its content is called Query By Image Content (QBIC), which allows users to retrieve images by analyzing visual features effectively [1]. In the retrieval process, the user provides a query image, which is analyzed according to specific criteria, and the system retrieves images with similar visual features [2, 3]. Developing an effective and efficient image feature descriptor has become a challenging task in image retrieval [4]. A scene or object may be captured under various conditions, such as different lighting sources, illustrations, and viewing angles. The image retrieval system offers an efficient way to search and browse for a desired image within a database of image sets [5]. High-level features are used to recognize images, while machines utilize low-level visual features. The difference between low-level features (e.g., pixels) and high-level

Received: October 26, 2024. Revised: November 14, 2024.

features (e.g., actions, objects, people) is known as the semantic gap [6, 7].

The semantic gap refers to the disconnect between raw data and meaningful interpretations, such as recognizing objects or actions in an image [8]. Initially, the low-level visual content is extracted from the query image in the retrieval system, and target images in the digital source are analyzed to construct feature descriptors/vectors. Suitable similarity measures are then calculated between the feature descriptor of the query image and each target image in the digital source [9]. Based on these calculated similarity measures, the target images are ranked to the query image [10]. Visual content such as texture, color, and shape plays a crucial role in producing desired outcomes based on user requirements [11, 12]. The efficient identification of the target object provides reliable and accurate feature descriptions for instance-level image retrieval, and low-dimensional representation results in faster retrieval efficiency and reduced storage costs [13]. Therefore, it is essential to develop an effective approach to recognizing target objects to create a discriminative and compact representation that enhances image retrieval [14, 15]. However, inadequate or inconsistent annotations can lead to poor retrieval performance, reducing the effectiveness of the image. To address this issue, the FCM-GBWSO approach is proposed for automatic image annotation and retrieval, enhancing the accuracy of image labeling and clustering. This approach improves retrieval performance by optimizing the annotation process, leading to more accurate and relevant search results.

The main contribution of this research is provided as follows:

- The FCM-GBWSO is employed to minimize the search space and increase retrieval efficiency. In WSO, GB is incorporated to solve premature convergence issues and avoid local optima.
- GBWSO is applied to determine the optimal number of clusters and centroids for FCM, enabling efficient image clustering. The GBWSO is optimized by utilizing the Euclidean distance between the cluster centroids and the query image.
- The features extracted from ResNet50 and color moments are transformed and selected using NCA, which helps convert the features to similar importance and select the best feature subset. This refined feature selection also enhances automatic annotation and retrieval efficiency.

This paper is summarized as follows: Section 2 provides a literature survey of existing methods. Section 3 illustrates an explanation of the proposed methodology. Section 4 involves experimental results, and the conclusion of this research paper is given in Section 5.

2. Literature survey

Hidayat [16] suggested hierarchical clustering on low-level features for image retrieval. The shape, texture, and color are the low-level features that were extracted and then hierarchically clustered. Then, the resultant clusters were validated to acquire an optimal number. In the retrieval progress, the query image features were extracted and compared with each feature's cluster centroid. The query outcome scores for all features were normalized, and these normalized scores were weighted to calculate an overall score, leading to improved performance. However, the hierarchical clustering suffered from poor discriminative power due to the inability to capture intricate image patterns and semantic data.

Mahalakshmi and Fatima [17] implemented an ensemble of Deep Learning (DL) methods to retrieve the images. At first, Convolutional Neural Network (CNN) based VGGNet-19 was applied to extract the feature and Euclidean distance-based similarity measures for image retrieval. Then, the Bidirectional-Long Short-Term Memory (Bi-LSTM) was utilized simultaneously to retrieve the textual documents. The implemented Bi-LSTM sequentially considered each word in a sentence which extracted the data and embedded it into a semantic vector. Nevertheless, the ensemble of DL struggled with high-dimensional feature fusion which led to inefficiencies in aligning and matching disparate modalities and resulted in suboptimal retrieval performance.

Keisham and Neelima [18] developed a Deep Neural Network-based deep Search and Rescue (DNN-SAR) for image retrieval. Initially, the Fast Average Peer Group (FAPG) filter was applied to eliminate the noise in the pre-processing phase. Then, the various features such as shape, color, and texture were extracted and then the feature vectors were computed. All these features were combined into a single feature by employing an average and weighted average model. Then, the combined features were clustered by utilizing a Sunflower Optimization Approach (SFO) which provides a better convergence rate. However, the DNN-SAR struggled in managing diverse scenes due to overfitting on specific training data which leads to reduced generalization.

Received: October 26, 2024. Revised: November 14, 2024.

Luo and Hu [19] presented an Adaptive Attention-ResNet (AA-ResNet) for image retrieval. The presented approach involved a feature extraction shown by the pattern, a complementary network, and a generator. The feature extraction defined by a pattern was utilized to extract the various feature levels which enabled the assessment of various aspects of the image. The design of the AA-ResNet effectively employed image data of the entire level and local level which enhanced the feature representation. Nevertheless, the AA-ResNet struggled with accurately capturing fine-grained information because it focused on the attention mechanism which reduced performance in distinguishing similar images with minor differences.

Jintanachaiwat and Siriborvornratanakul [20] introduced a ResNet50 for image search similarity. Global Max Pooling (GAP) were used to extract the features and the metric embedding layer formed the last embeddings with a dimension. ResNet50 employed skip connections to prevent vanishing gradient and enabled the training of deep networks. However, ResNet50 struggled with capturing high-level semantic data due to its fixed architecture and relatively shallow depth which resulted in less effective retrieval performance.

In the overall analysis, the existing methods had limitations like the inability to capture intricate image patterns, inadequate or inconsistent annotations, struggle to manage diverse scenes and capture high-level semantic data. To address these limitations, the FCM-GBWSO is proposed for image annotation and retrieval by enhancing the ability to capture image patterns and high-level semantic data via enhanced clustering and optimization process. It ensures more consistent and accurate annotations which manage diverse scenes effectively. This leads to better retrieval performance and more reliable effectiveness.

3. Proposed methodology

This research proposes FCM-GBWSO for automatic image annotation and retrieval process. It contains two difference process performed in FCM-GBWSO like online process (querying) and an offline process (indexing). The significant process of the offline procedure is a dataset, pre-processing, extraction, transformation, selection, and SSR by applying FCM-GBWSO. Similarly, the query image is indexed effectively and then similarity matching is performed with a subset of images for retrieving an appropriate image. Fig. 1 depicts a block diagram for the FCM-GBWSO method.

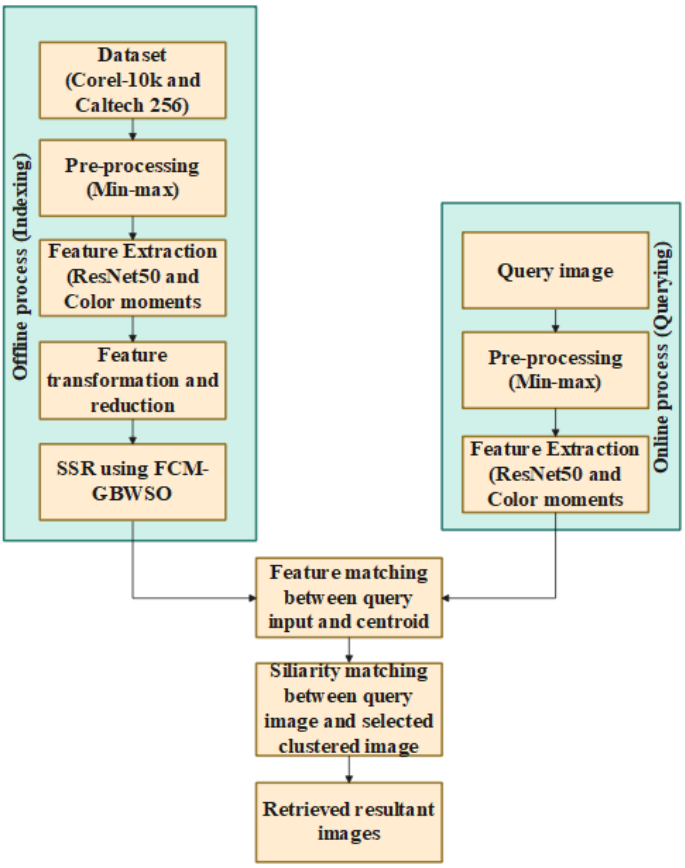

Figure. 1 Block diagram for the FCM-GBWSO method

Received: October 26, 2024. Revised: November 14, 2024.

Figure. 2 Sample images of Corel-10k dataset

Figure. 3 Sample images of the Caltech 256 dataset

Figure. 4 Sample images of the Corel 1k dataset

3.1 Dataset

In this research, the Corel-10k [21], Caltech 256 [22], Corel 1k [23], Corel 1.5k [18], and Corel 5k [24] are used to determine the proposed approach performance which is explained as follows:

Corel-10k: The Corel-10k contains 100 categories of images like aviation, dog, art, cat, tiger, owl, lion, etc. Each category contains 100 images with an overall of 10,000 images. Also, it involves a digitally zoomed version of different-sized images. Corel-10k contains 80% of training and 20% of testing respectively. Fig. 2 represents a sample image of the Corel-10k dataset.

Caltech 256: It includes 30,607 images which is divided into 256 categories. Of all total images, 80% are used for training whereas the remaining 20% are employed for the testing process. Each image belongs to 101 various categories and the number of images in every category range from 80 to 827 correspondingly. Fig. 3 indicates sample images of the Caltech 256 dataset. Then, the obtained images are passed through the pre-processing step.

Corel 1k: It has 1000 images and divides into 10 classes. Overall, it contains 100 images which are present in each class with 256 X 384 size. The 70% of images are selected as training images and remaining 30% are chosen
as testing images in the proposed approach. Figure 4 indicates a sample images of the Corel 1k dataset.

Corel 1.5k: It contains overall 1500 images that are split into 15 classes and 100 images are present in each class with a 256 X 384 size. Among the overall 1500 images, 80% of images are chosen for training and 20% of images are considered for testing purpose. Figure 5 represents a sample image of Corel 1.5k dataset.

Figure. 5 Sample images of Corel 1.5k dataset

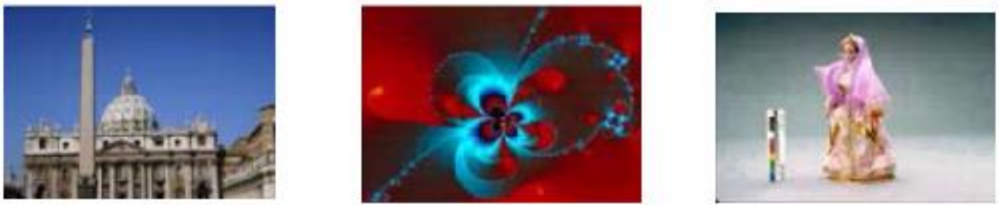

Figure. 6 Sample images of Corel 5k dataset

Corel 5k: It has 5000 images which is divided into 50 groups and 100 images are present in each class with an equal size of 256 X 384 size. Among overall 5000 images, 70% of images are considered for training and 30% for testing process respectively. Figure 6 denotes a sample images of Corel 5k dataset.

3.2 Pre-processing

Then, the gathered images are fed into pre-processing using the normalization model. Min-max [25] method is preferred over z-score when the data has no outliers and it scales pixel values of images to a fixed range [0,1] which preserves the relationships when removing distortions caused by varying intensities. An input intensity pixel is enhanced by altering pixel limits by applying min-max which is formulated in Eq. (1).

$$l' = (I - \min)\frac{newmax - newmin}{max - min} + newmin \quad (1)$$

Where I represents input image, min and max indicates minimum and maximum input image value, l' determines normalized image, whereas $newmin$ and $newmax$ depicts value of new intensity l'. This enhances the image annotation and retrieval performance by enhancing the feature consistencies across images resulting in more accurate performance. Then, the normalized image is fed into the feature extraction process.

3.3 Feature extraction

A pre-processing image are provided as input to various types of techniques like ResNet50 and Color moments. Through image retrieval, feature extraction is a progress of initial reduction and changing images from pixel to feature space. Detailed information is explained as follows:

ResNet50: ResNet50 [26] is a variant of Residual Network which is established for training the Deep Neural Networks (DNN) with 50 layers. It constructs the residual learning concept which solves the vanishing gradient issue which allows the network to learn rapidly and more efficiently even as the network depth is enhanced. This is achieved via skip connections or shortcut connections which allow the gradient to be backpropagated directly to previous layers. ResNet50 is considered for its depth and complex design components which are specifically designed to enhance gradient propagation and a rich feature extraction. It contains a convolutional layer, Batch Normalization (BN), and Rectified Linear Unit (ReLU) activation function layers which process together to enhance feature extraction and improve training stability by normalizing inputs. In these initial layers, the model detects basic features like textures, edges, corners, shapes, and color patterns.

However, as the network depth increases, it prioritizes features such as shapes which leads to reduced significance of color in later stages. Hence, color moments are used in feature extraction which is described below.

Color Moments: Color moments are an effective and simple method to indicate a color feature and are dispersed in low-order moments that is mean, variance, and skewness utilized for determining image's color dissemination. Hence, color features are extracted via the mean (E_i), skewness (SW_i), and standard deviation (σ_i) which is formulated in Eqs. (2) to (4).

$$E_i = \frac{1}{N}\sum_{j=1}^{N} P_{i,j} \quad (2)$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}(P_{i,j} - E_i)^2} \quad (3)$$

$$SW_i = \sqrt[3]{\frac{1}{N}\sum_{j=1}^{N}(P_{i,j} - E_i)^3} \quad (4)$$

Where $P_{i,j}$ indicates color elements i of j^{th} pixel and overall image pixel is denoted as N. Color moments effectively capture the color distribution in an image by utilizing statistical moments from each color channel. This simplifies the comparison among images while managing significant color information. Color moments are efficient for color-based search and retrieval which provides robustness to minor variations like noise or lightning changes. ResNet50 and color moments are integrated to establish and total Feature Vector (FV) which is expressed in Eq. (5).

$$FV = \{ResNet, E_i, \sigma_i, SW_i\} \quad (5)$$

The concatenated features are provided as input for further process to normalize and select finest features.

3.4 Feature transformation and selection

A concatenated FV is fed as input for the transformation of features and every feature contains a range value. A feature with a greater feature value manages and assigns higher weight than fewer feature values in similarity matching. Hence, the feature transformation is significant to convert every feature with alike significance. An intra-normalization is applied where every feature vector is normalized using Eq. (6).

$$FV_{i,j} = \frac{FV_{i,j} - Fmax_i}{Fmax_i - Fmin_i} \quad (6)$$

Where, $Fmax_i$ and $Fmin_i$ represents the maximum and minimum values of all feature vectors correspondingly. The transformed features employ Neighborhood Component Analysis (NCA) to choose an optimum feature set by removing the inappropriate features. NCA is used to categorize multi-variate data into various class depending on distance parameters among data. It is non-parametric and does not need any parameter or assumption regarding the sample's statistical data. In NCA, feature ranking is performed with regularization to learn the feature weight for minimization of the objective function. Hence, the weight value indicates each feature's contribution to linear transformation's discriminative effectiveness obtained by NCA. Based on the weights, the features are ranked while the higher weight represents more essential features for processing retrieval functions. At last, the top-ranked features are selected with a length for further processing.

3.5 Search Space Reduction (SSR)

During feature selection, the retrieval efficiency is affected while significant features are discarded. The chosen features are provided as input to SSR which is necessary to increase retrieval accuracy. In the query phase, the traditional image retrieval matching for large data repositories impacts the retrieval performance. Hence, it is vital to select an initial congregation of extremely appropriate images from the repository. The FCM [27] model is an effective clustering method which is an unsupervised approach used to select the initial image subset. Here, search space is minimized by integrating FCM with GBWSO. The FCM needed a cluster member pre-identification and it is struggled due to a dead unit problem that indicates an insufficient cluster center initialization. Hence, the above problem is addressed by integrating the GBWSO with FCM. From the GBWSO, the cluster center's optimal number is determined and that is evaluated to cluster number for FCM. In the subsequent section, the GBWSO and FCM are explained in detail.

3.5.1. GBWSO

FCM faces challenges due to insufficient initialization of cluster center initialization, therefore GBWSO is utilized to determine the optimal number of the cluster center. A random k value and cluster centroid are passed to GBWSO to acquire the best cluster number. WSO [28] is a population-based approach such as numerous metaheuristic algorithms. The candidate solutions to an optimization issue with an d dimensional space and n population size (white shark) is formulated in Eq. (7).

$$w = \begin{bmatrix} w_1^1 & w_2^1 & \dots & w_d^1 \\ w_1^2 & w_2^2 & \dots & w_d^2 \\ \cdots & \cdots & \dots & \cdots \\ w_1^n & w_2^n & \dots & w_d^n \end{bmatrix} \quad (7)$$

Where w indicates the location of every shark in search domain and d represents a number of variables for the provided task.

Speed of Movement towards Prey: A white shark identifies a location of prey by hearing a pause in waves by moving prey which is expressed in Eq. (8).

$$u_{k+1}^i = \mu\left[u_k^i + p_1(w_{gbest_k} - w_k^i) \times c_1 + p_2(w_{best}^{v_k^i} - w_k^i) \times c_2\right] \quad (8)$$

$i = 1,2,\dots,n =$ size index, u_{k+1}^i denotes speed vector of i^{th} shark at $k+1$ iteration, μ indicates scaling factor, u_k^i represents current position of i^{th} shark at k iteration, and v^i represents shark index vector obtaining the best location using Eq. (9)

$$v^i = [n \times rand(1,n)] + 1 \quad (9)$$

Where $rand(1,n)$ denotes randomly produced numbers which have a domain distribution [0,1].

$$p_1 = p_{max} + (p_{max} - p_{min}) \times e^{-(\frac{4m}{k})^2} \quad (10)$$

$$p_2 = p_{min} + (p_{max} - p_{min}) \times e^{-(\frac{4m}{k})^2} \quad (11)$$

Where k represents present, m denotes maximum iterations, and p_{min} and p_{max} depicts initial and subordinate velocity for the motion of a white shark using Eq. (10) and (11). Then, those values are determined as 0.5 and 1.5 using Eq. (12).

$$\mu = \frac{2}{|2-\tau-\sqrt{\tau^2-4\tau}|} \quad (12)$$

Where τ denotes accelerating factor.

Optimal Prey Movement Direction: The update position strategy represented in Eq. (13) is employed to determine the white shark behavior as they move towards prey.

$$w_{k+1}^i = \begin{cases} w_k^i \rightarrow \oplus\, w_0 + u.a + l.b; rand < mv \\ w_k^i + \frac{u_k^i}{f}; rand \geq mv \end{cases} \quad (13)$$

Where $\oplus$ represents the bitwise operation outcomes. Eqs. (14-16) are described a and b as binary vectors.

$$a = sgn(w_k^i - u) > 0 \quad (14)$$

$$b = sgn(w_k^i - 1) < 0 \quad (15)$$

$$w_0 = \oplus\, (a, b) \quad (16)$$

Where u indicates undefined parameter in the current context, sgn determines sign function, and w_0 represents combined state based on conditions a and b. The white shark's wavy motion frequency and shark attack time multitude are expressed in Eqs. (17) and (18).

$$f = f_{min} + \frac{f_{max} - f_{min}}{f_{max} + f_{min}} \quad (17)$$

$$mv = \frac{1}{(a_0 + e^{\frac{\frac{k}{2} - k}{a_1}})} \quad (18)$$

Where a_0 and a_1 represents location constants that are utilized to manage exploitation and exploration and mv indicates shark attack time multitude.

Optimal Shark Movement Direction: Shark maintains its place who is closer to the target which is formulated using Eq. (19).

$$w_{k=1}'^{i} = w_{gbestk} + r_1 \overrightarrow{D_{wsgn}}(r_2 - 0.5) r_3 < S_s \quad (19)$$

Where $w_{k=1}'^{i}$ represents upgraded shark's location, $sgn(r_2 - 0.5)$ becomes 1 or -1 for modifying the search path, r_1, r_2, and r_3 indicates random number, $\overrightarrow{D_w}$ denotes both shark and target length, using Eq. (20), and S_s depicts parameters for reflecting the white shark power by applying Eq. (20).

$$\overrightarrow{D_w} = |rand \times (w_{gbest} - w_k^i)| \quad (20)$$

$$S_s = |1 - e^{\frac{a_2 \times k}{k}}| \quad (21)$$

Where a_2 indicates the location factor for regulating exploitation and exploration. However, the variety of WSO approaches diminishes during the later stage of determining WSO which results in potential problems with accuracy and convergence speed. The GB mechanism provides the selection of the most appropriate direction for white sharks and makes them gradually progress near the best solutions which prevents premature convergence to avoid local optima issues. Therefore, GB is incorporated into the WSO model to increase the population diversity after updating each search agent position. This manages a balance among the algorithm's local exploitation and its ability for global search which leads to enhanced convergence speed. The GB mechanism is resultant of Bare-Bones Particle Swarm Optimization (BBPSO). In this mechanism, the parameter R is used for guiding each individual. If the probability of random process is less than R , the Gaussian distribution updates the individual position in the following evaluation. The mathematical formula for the GB mechanism is expressed in Eq. (22).

$$w_{k,GB}^j = \begin{cases} G(\frac{(w_{gbest,k-1} + w_k^j)}{2}, |w_{gbest,k-1} - w_k^j|) & if\ rand < R \\ w_k^{j1} + r_4(w_k^{j2} - w_k^{j3}) & otherwise \end{cases} \quad (22)$$

Where $w_{k,GB}^j$ indicates j^{th} white shark's new position by utilizing a GB; $w_{gbest,k-1}$ represents global best solution acquired in $k-1$ iterations, r_4 denotes random number in [0,1] interval, G depicts Gaussian distribution with mean $\frac{(w_{gbest,k-1} + w_k^j)}{2}$ and standard deviation $(w_{gbest,k-1} + w_k^j)$. The GB-WSO increases the exploration and exploitation capabilities and assists the algorithm to escape local optima and better explore search space. This integration enhances convergence speed and solution accuracy.

3.5.2. FCM-GBWSO

The FCM-GBWSO is the combination of the GBWSO and FCM model. The GBWSO is utilized to identify the cluster number and centroid that is provided as the initial seed value of FCM. The FCM-GBWSO process is explained as follows,

- The white shark is initialized by centroids from image FV and k random values. A single prey position determines cluster centroid in the clustering background.
- For each white shark, the fitness function is calculated by utilizing a clustering criterion i.e., distance. The white shark location is updated depending on the prey. The same measures are performed till the white shark spreads maximum iterations.
- Establish GBWSO based on optimal results from GBWSO and it generates the cluster centroids and clustered images.
- At last, cluster centroid and query image are determined together, and then the cluster with fewer distances is measured for retrieval progress.

3.6 Retrieval stage

In this phase, the search utilized a predefined query image in image retrieval is determined in FCM-GBWSO. For a provided query image, a similar feature set is acquired and generated utilizing feature extraction, transformation, and selection progress. Then, the query feature image is determined with cluster centroid acquired from FCM-GBWSO. At last, cluster images that are greatly matched with query images are retrieved from the data. Figs. 7 and 8 show a query image and retrieved images for Corel-10k and Caltech 256 datasets.

Figure. 7 Query and retrieved images for Corel-10k dataset

Figure. 8 Query and retrieved images for Caltech 256 dataset

4. Experimental results

The proposed FCM-GBWSO is simulated based on MATLAB R2018a with Windows 7 operating system, an Intel i5 processor, and 6 GB RAM. The performance measures like average precision, average recall, and average f1-score are used to determine the proposed approach performance using Eqs. (23) to (25).

$$Average\ Precision = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FP_i} \tag{23}$$

$$Average\ Recall = \frac{1}{N}\sum_{i=1}^{N}\frac{TP_i}{TP_i+FN_i} \tag{24}$$

$$Average\ F1-score = \frac{1}{N}\sum_{i=1}^{N} 2 \times \frac{Precision_i \times Recall_i}{Precision_i + Recall_i} \tag{25}$$

Where N represents total number of classes, TP_i indicates True Positive for class i, FP_i denotes False Positive for class i, TN_i determines True Negative for class i, and FN_i illustrates False Positive for class i.

4.1 Performance analysis

Table 1 indicates a performance analysis of different feature extraction methods. The existing methods like VGG16, ResNet50, and color moments are compared with ResNet50+Color moments. The ResNet50+Color moments achieve a better average precision of 0.98 and 0.96 using Corel-10k and Caltech 256 datasets due to effective extraction with robust color descriptors which enhances both texture and color information. This approach captures richer visual information and leads to better outcomes. Also, integrating the benefits of ResNet50's depth assists in achieving better generalization.

Table 2 represents the performance analysis of different clustering methods. The existing methods like K-means clustering with GBWSO, Hierarchical clustering with GBWSO, and Density-Based Spatial Clustering of Applications with Noise-based GBWSO (DBSCAN-GBWSO) are compared with proposed FCM-GBWSO. The proposed FCM-GBWSO achieves a better average precision of 0.98 and 0.96 using Corel-10k and Caltech 256 datasets due to FCM's flexibility in assigning data points to multiple clusters with varying membership degrees. This enables FCM to capture more nuanced patterns and makes it efficient in managing overlapping clusters.

Table 3 demonstrates a performance evaluation of different optimization methods. The proposed FCM-GBWSO achieves a better average precision of 0.98 and 0.96 using Corel-10k and Caltech 256 datasets compared to the FCM-Whale Optimization Approach (WOA), FCM-Golf Optimization Approach (WOA), and FCM-WSO. Due inclusion of GB in WSO increases global search ability and prevents premature convergence. This combination enables more refined optimization and effectively balances the exploration-exploitation phase which leads to high performance.

Table 1. Different feature extraction methods

Methods	Dataset	Average Precision	Average Recall	Average F1-score
VGG16	Corel-10k	0.82	0.79	0.80
ResNet 50		0.86	0.81	0.76
Color moments		0.89	0.84	0.78
ResNet 50 + Color moments		**0.98**	**0.94**	**0.95**
VGG16	Caltech 256	0.79	0.75	0.78
ResNet 50		0.82	0.79	0.82
Color moments		0.87	0.83	0.85
ResNet 50 + Color moments		**0.96**	**8.652**	**0.93**

Table 2. Different clustering methods

Methods	Dataset	Average Precision	Average Recall	Average F1-score
K-GBWSO	Corel-10k	0.84	0.81	0.82
H-GBWSO		0.89	0.87	0.89
DBSCAN-GBWSO		0.92	0.90	0.90
Proposed FCM-GBWSO		0.98	0.94	0.95
K-GBWSO	Caltech 256	0.87	0.89	0.81
H-GBWSO		0.92	0.91	0.84
DBSCAN-GBWSO		0.94	0.92	0.89
Proposed FCM-GBWSO		**0.96**	**8.652**	**0.93**

Received: October 26, 2024. Revised: November 14, 2024.

Table 3. Different optimization methods

Methods	Dataset	Average Precision	Average Recall	Average F1-score
FCM-WOA	Corel-10k	0.86	0.90	0.85
FCM-GOA		0.92	0.89	0.89
FCM-WSO		0.95	0.92	0.91
Proposed FCM-GBWSO		0.98	0.94	0.95
FCM-WOA	Caltech 256	0.90	6.893	0.83
FCM-GOA		0.92	7.308	0.89
FCM-WSO		0.94	7.983	0.92
Proposed FCM-GBWSO		**0.96**	**8.652**	**0.93**

4.2 Comparative analysis

Table 4 determines a comparative analysis of existing methods. Table 5 represents a comparative analysis of existing methods using MAP. The existing methods like Hierarchical clustering [16], Ensemble DL methods [17], DNN-SAR [18], and ResNet50 [20] are compared with the proposed FCM-GBWSO. The proposed FCM-GBWSO achieves a better average precision of 0.98, 0.96, and 0.97 using Corel-10k, Caltech 256, and Corel 1k datasets. The proposed FCM-GBWSO achieves a better Mean Average Precision (MAP) of 97.67%, 94.20%, and 97.46%, and 93.47% compared to existing method [18] using Corel 1k, Corel 1.5k, Corel 5k, and Caltech 256 datasets due to it provides an enhanced clustering performance by incorporating FCM's flexible membership with GBWSO which leads to better convergence and avoids local optima.

Table 4. Comparative analysis of existing methods

Methods	Dataset	Average Precision	Average Recall
Hierarchical clustering [16]	Corel-10k	0.62	0.11
Ensemble DL methods [17]		0.94	0.88
Proposed FCM-GBWSO		0.98	0.94
DNN-SAR [18]	Caltech 256	N/A	7.881
EfficientNet B7 [20]		0.93	N/A
Proposed FCM-GBWSO		0.96	8.652
Hierarchical clustering [16]	Corel 1k	0.81	0.16
EfficientNet B7 [20]		0.94	N/A
Proposed FCM-GBWSO		0.97	0.25

Table 5. Comparative analysis of existing methods using MAP

Methods	Datasets	MAP (%)
DNN-SAR [18]	Corel 1k	93.91
Proposed FCM-GBWSO		97.67
DNN-SAR [18]	Corel 1.5k	90.33
Proposed FCM-GBWSO		94.20
DNN-SAR [18]	Corel 5k	95.4
Proposed FCM-GBWSO		97.46
DNN-SAR [18]	Caltech 256	86.13
Proposed FCM-GBWSO		93.47

4.3 Discussion

The advantages of the proposed FCM-GBWSO method, along with the limitations of existing approaches, are discussed in this section. Existing methods, such as hierarchical clustering [16], suffer from poor discriminative power due to their inability to capture intricate image patterns and semantic data. Ensemble DL [17] faces challenges with high-dimensional feature fusion, leading to inefficiencies in aligning and matching disparate modalities. DNN-SAR [18] struggles to manage diverse scenes due to overfitting on specific training data, which results in reduced generalization. ResNet50 [20] encounters difficulties in capturing high-level semantic data because of its fixed architecture and relatively shallow depth, which in turn affects retrieval performance. The proposed FCM-GBWSO overcomes these limitations. FCM effectively manages overlapping data clusters and enhances feature extraction performance, while GB increases noise resistance and cluster sharpness. WSO optimizes the clustering process by balancing exploitation and exploration, leading to faster convergence and more accurate retrieval. This improves both the retrieval speed and the relevance of retrieved images.

Received: October 26, 2024. Revised: November 14, 2024.

5. Conclusion

This research proposes FCM-GBWSO for the automatic image annotation and retrieval process. FCM-GBWSO is utilized to generate optimal clustering in images to achieve efficient retrieval by using a query image from an online process. Hence, FCM-GBWSO efficiently retrieves images with similar content from the Corel-10k and Caltech 256 datasets. An image from the offline process is enhanced through normalization, and relevant features are extracted using ResNet50 and color moments. NCA is used for feature transformation and selection to normalize and choose optimal feature subsets. From the overall analysis, it is concluded that FCM-GBWSO achieves better average precision of 0.98, 0.96, and 0.97 using Corel-10k, Caltech 256, and Corel 1k datasets respectively, compared to existing methods like Hierarchical Clustering and DNN-SAR. In the future, different clustering approaches will be considered to further enhance the model's performance.

Notation Description

Symbols	Description
I	Input image
min and max	Minimum and maximum input image value
l'	Normalized image
$newmin$ and $newmax$	Value of new intensity l'
$P_{i,j}$	Color elements i of j^{th} pixel
N	Overall image pixel
(E_i)	Mean
(SW_i)	Skewness
(σ_i)	Standard deviation
$Fmax_i$ and $Fmin_i$	Maximum and minimum values of all feature vectors
w	Location of every shark in search domain
d	Number of variables for the provided task
$rand(1,n)$	Randomly produced numbers which have a domain distribution [0,1]
u_{k+1}^i	Speed vector of i^{th}shark at $k+1$ iteration
μ	Scaling factor
u_k^i	Current position of i^{th} shark at k iteration
v^i	Shark index vector obtaining the best location
m	Maximum iterations
p_{min} and p_{max}	Initial and subordinate velocity for the motion of a white shark
τ	Accelerating factor
$\oplus$	Bitwise operation outcomes
a and b	Binary vectors
u	Undefined parameter in the current context
sgn	Sign function
w_0	Combined state based on conditions a and b
a_0 and a_1	Location constants that are utilized to manage exploitation and exploration
mv	Shark attack time multitude
$w'^{i}_{k=1}$	Upgraded shark's location
r_1, r_2, r_3, and r_4	Random number
$\overrightarrow{D_w}$	Shark and target length
S_s	Parameters for reflecting the white shark power
$w_{k,GB}^j$	j^{th} white shark's new position by utilizing a GB
$w_{gbest,k-1}$	Global best solution acquired in $k-1$ iterations
G	Gaussian distribution with mean $\frac{\left(w_{gbest,k-1}+w_k^j\right)}{2}$ and standard deviation $\left(w_{gbest,k-1}+w_k^j\right)$

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, done by 5th and 6th author, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 3rd and 4th author. The supervision and project administration, have been done by 1st and 2nd author.

References

[1] S. Bakheet, A. Al-Hamadi, E. Soliman, and M. Heshmat, "Hybrid bag-of-visual-words and FeatureWiz selection for content-based visual information retrieval", *Sensors*, Vol. 23, No. 3, p. 1653, 2023.

[2] B. Patel, B.M. Singh, and K. Yadav, "DDBTC approach with binary particle swarm optimization for greedy-DCNN based CBIR system", *Biomedical Signal Processing and Control*, Vol. 76, p. 103710, 2022.

[3] S. Sharma, V. Gupta, and M. Juneja, "A novel unsupervised multiple feature hashing for image retrieval and indexing (MFHIRI)", *Journal of*

Visual Communication and Image Representation, Vol. 84, p. 103467, 2022.

[4] Y. Wang, F. Wang, F. Liu, and X. Wang, “Securing content-based image retrieval on the cloud using generative models”, *Multimedia Tools and Applications*, Vol. 81, No. 22, pp. 31219-31243, 2022.

[5] Z. Zeng, J. Wang, B. Chen, T. Dai, S.T. Xia, and Z. Wang, “Pyramid hybrid pooling quantization for efficient fine-grained image retrieval”, *Pattern Recognition Letters*, Vol. 178, pp. 106-114, 2024.

[6] M.S. Sayed, A.A. Gad-Elrab, K.A. Fathy, and K.R. Raslan, “Unsupervised Content Based Image Retrieval Using Pre-Trained CNN and PCNN Features Extractors”, *International Journal of Intelligent Engineering & Systems*, Vol. 16, No. 1, pp. 584-596, 2023.

[7] W.C. Lin, “Block-based pseudo-relevance feedback for image retrieval”, *Journal of Experimental & Theoretical Artificial Intelligence*, Vol. 34, No. 5, pp. 891-903, 2022.

[8] H. Rastegar, and D. Giveki, “Designing a new deep convolutional neural network for content-based image retrieval with relevance feedback”, *Computers and Electrical Engineering*, Vol. 106, p. 108593, 2023.

[9] N. Varish, “A modified similarity measurement for image retrieval scheme using fusion of color, texture and shape moments”, *Multimedia Tools and Applications*, Vol. 81, No. 15, pp. 20373-20405, 2022.

[10] D. Singh, J. Mathew, M. Agarwal, and M. Govind, “DLIRIR: Deep learning based improved reverse image retrieval”, *Engineering Applications of Artificial Intelligence*, Vol. 126, Part A, p. 106833, 2023.

[11] R.A. Narayanaswamy, V.K. Gowda, and G.N.K. Murthy, “Content-based Image Retrieval Using the 2-Dimensional Convolutional Neural Network”, *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 4, pp. 67-76, 2024, doi: 10.22266/ijies2024.0831.06.

[12] S. Kumar, J. Pradhan, A.K. Pal, S.K.H. Islam, and M.K. Khan, “Radiological image retrieval technique using multi-resolution texture and shape features”, *Multimedia Tools and Applications*, Vol. 81, No. 10, pp. 13633-13660, 2022.

[13] B.J. Zhang, G.H. Liu, and Z. Li, “Image retrieval using unsupervised prompt learning and regional attention”, *Expert Systems with Applications*, Vol. 247, p. 122913, 2024.

[14] F.A.A. Salih, and A.A. Abdulla, “Two-layer content-based image retrieval technique for improving effectiveness”, *Multimedia Tools and Applications*, Vol. 82, No. 20, pp. 31423-31444, 2023.

[15] S. Chavda, and M. Goyani, “Robust image retrieval using CCV, GCH, and MS-LBP descriptors”, *Multimedia Tools and Applications*, Vol. 81, No. 3, pp. 4039-4072, 2022.

[16] R. Hidayat, A. Harjoko, and A. Musdholifah, “A Robust Image Retrieval Method Using Multi-Hierarchical Agglomerative Clustering and Davis-Bouldin Index”, *International Journal of Intelligent Engineering & Systems*, Vol. 15, No. 2, pp. 441-453, 2022, doi: 10.22266/ijies2022.0430.40.

[17] P. Mahalakshmi, and N.S. Fatima, “Ensembling of text and images using deep convolutional neural networks for intelligent information retrieval”, *Wireless Personal Communications*, Vol. 127, No. 1, pp. 235-253, 2022.

[18] N. Keisham, and A. Neelima, “Efficient content-based image retrieval using deep search and rescue algorithm”, *Soft Computing*, Vol. 26, No. 4, pp. 1597-1616, 2022.

[19] J. Luo, and D. Hu, “An image classification method based on adaptive attention mechanism and feature extraction network”, *Computational Intelligence and Neuroscience*, Vol. 2023, No. 1, p. 4305594, 2023.

[20] W. Jintanachaiwat, and T. Siriborvornratanakul, “Vision-based image similarity measurement for image search similarity”, *International Journal of Information Technology*, Vol. 15, No. 8, pp. 4125-4130, 2023.

[21] Corel-10k dataset link: https://www.kaggle.com/datasets/michelwilson/corel10k

[22] Caltech 256 dataset link: https://www.kaggle.com/datasets/jessicali9530/caltech256

[23] Corel 1k dataset link: https://www.kaggle.com/datasets/elkamel/corel-images

[24] Corel 5k dataset link: https://www.kaggle.com/datasets/parhamsalar/corel5k

[25] Y.K. Saheed, A.I. Abiodun, S. Misra, M.K. Holone, and R. Colomo-Palacios, “A machine learning-based intrusion detection for detecting internet of things network attacks”, *Alexandria Engineering Journal*, Vol. 61, No. 12, pp. 9395-9409, 2022.

[26] L. Zhang, Y. Bian, P. Jiang, and F. Zhang, “A transfer residual neural network based on

ResNet-50 for detection of steel surface defects", *Applied Sciences*, Vol. 13, No. 9, p. 5260, 2023.

[27] S. Tokat, K. Karagul, Y. Sahin, and E. Aydemir, "Fuzzy c-means clustering-based key performance indicator design for warehouse loading operations", *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 8B, pp. 6377-6384, 2022.

[28] N.A. Kadhim, A.A. Obed, A.J. Abid, H. Kotb, and A. Emara, "Optimal PV Reconfiguration Under Partial Shading Based on White Shark Optimization", *IEEE Access*, Vol. 12, pp. 27385-27398, 2024.