

Content-Based Image Retrieval Using Hybrid Densenet121-Bilstm and Harris Hawks Optimization Algorithm

Sanjeevaiah K., Hyderabad Institute of Technology and Management, India

Tatireddy Subba Reddy, Department of Computer Science and Engineering, B V Raju Institute of Technology, Narsapur, India*

Sajja Karthik, R.V.R. & J.C. College of Engineering, India

Mahesh Kumar, Mahatma Gandhi Institute of Technology, India

Vivek D., B.V. Raju Institute of Technology, India

ABSTRACT

In the field of digital data management, content-based image retrieval (CBIR) has become one of the most important research areas, and it is used in many fields. This system searches a database of images to retrieve most visually comparable photos to a query image. It is based on features derived directly from the image data, rather than on keywords or annotations. Currently, deep learning approaches have demonstrated a strong interest in picture recognition, particularly in extracting information about the features of the image. Therefore, a Densenet-121 is employed in this work to extract high-level and deep characteristics from the images. Afterwards, the training images are retrieved from the dataset and compared to the query image using a Bidirectional LSTM (BiLSTM) classifier to obtain the relevant images. The investigations are conducted using a publicly available dataset named Corel, and the f-measure, recall, and precision metrics are used for performance assessment. Investigation outcomes show that the proposed technique outperforms the existing image retrieval techniques.

KEYWORDS

BiLSTM, CBIR, Deep Learning, Densenet-121, Feature Extraction

Introduction

The rapid proliferation of digital images in cyberspace has inspired research into developing efficient image content management system. In the recent decade, the rapid advancement of digital capturing devices and social media has led to a huge development of image database systems (Desai et al., 2021; Zhong et al., 2021; Desai et al., 2021). Acquiring important data from these enormous databases has stimulated the research community's desire to find effective alternatives that do not rely on textual descriptions for each image. This investigation led to the discovery of a technique for content-based picture retrieval (CBIR).

In many areas, the CBIR is utilized like crime preventing, video processing, digital albums, biodiversity, medical imaging, and other areas which need image recognition. In CBIR, the images are automatically indexed based on its low level features extracted from the image like shape and color which are only used for image retrieval (Keisham et al., 2022; Kumar et al., 2022; Chen et al., 2021). The two most critical aspects of the CBIR program's implementation are feature representation and similarity measurement.

DOI: 10.4018/IJSI.315661

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Color-based features, texture characteristics, and shape features are among the low-level features examined for clustering and retrieving images (Putzu et al., 2020). The value of pixels in RGB, HSV, and other colour spaces is extracted and represented as the image's feature vector in color-based features (Sundararajan et al., 2019; Hu et al., 2021; Zhang et al., 2022; Sezavar et al., 2019). The content of an image cannot be fully described by colour alone. As a result, employing colour attributes individually could not reach adequate accuracy. Therefore, to compare images, texture features are necessary.

In an image, texture is one of the crucial surface property which represents the more important surface related features like clouds, tiles, bricks etc. it is described by the similarity of visual patterns. The shape identifiers do not indicate that the complete shape of the image is described, but rather that the shape of a specific section of the image is described. Shapes are also utilised for segmentation and contour detection (Sharif et al., 2019; Öztürk et al., 2020; Ashraf et al., 2020). Scaling, rotation, and Invariance for translation are the approaches utilised for the shape descriptor. Using the discrete cosine transform, wavelet transform, discrete Fourier transform or a combination of them, some relevant texture features can be retrieved from images (Bibi et al., 2020).

Retrieval systems nowadays typically use a combination of texture, shape and colour features of images to attain high accuracy. Feature extraction approaches such as color auto correlogram, color correlogram, color histogram and inter-channel voting among saturation and hue can be utilised to obtain colour features of an image (Khalid et al., 2020; Srivastava et al., 2019; Hidayat et al., 2019). GLCM, LTrP, LDP, LEP CS LBP, and ULBP can all be used to extract texture features. Furthermore, several picture retrieval methods have been previously created. Unfortunately, none of the aforementioned techniques are very efficient for huge image datasets. Moreover, to increase retrieval performance, better features must be extracted. While it may be possible to extract more complicated low-level characteristics from images, increasing the calculation time would increase the size of the feature vector and slow down retrieval speed (Pradhan et al., 2020; Alshehri et al., 2020). Deep learning is the important concept that has been shown to eliminate the semantic gap. The deep learning architectures are very effective which learns the features from various levels of data.

Therefore in this paper an effective deep learning based technique called Densenet-121 is implemented to extract the important features of the image. Following feature extraction, an efficient method is required to search through a large number of features and accomplish the retrieval operation. This search is effectively performed by Bidirectional Long Short Term Memory (BiLSTM) and the hyperparameters are optimized by Harris hawks optimization algorithm. The main contributions of the works are listed as below.

A novel hybrid Densenet121-BiLSTM recognition model is introduced for content based image retrieval. Model performance improvement with the help of Harris hawks optimization algorithm.

To evaluate the system efficiency, several experiments are conducted on Corel dataset with various performance metrics.

The remaining portions of this work are structured as follows: The second section provides a quick overview of previous research efforts. Section 3 explains the concepts of the approaches and procedures needed to construct this framework. Section 4 explains the testing procedure and results, as well as providing brief explanations. Section 5 finishes with an overview of the research as well as suggestions for further research.

Related Works

Khan et al. (Khan et al., 2021) proposed an image retrieval mechanism based on support vector machine classifier and Genetic Algorithm (GA) as a feature descriptor. For feature extraction, they used Bi-Orthogonal wavelets, Daubechies Wavelet and Haar Wavelet for initial three color components. Then the important features were selected by GA which are given to the SVM for classification.

The similarity measurement among retrieved images and query image was conducted by L2 Norm function. To evaluate the performance of the technique 25 existing techniques were used on four datasets named kvasir, CIFAR-10, Oxford Flower, and WANG. During the evaluation process, the authors effectively handled the class imbalance problem.

Pathak and Raju (Pathak et al., 2021) performed feature fusion for CBIR framework. To extract the image features, the hybrid framework called Inception-Darknet-53 was introduced. In this approach, the layers of inception network was combined with Darknet-53 architecture. One GN layer was utilized in the hybrid architecture to stabilize the learning process and decrease the no of training epochs. Besides, the handcraft features were extracted by HIS and RGB color space. HOG based technique was implemented by the authors to extract the shape features. For performance assessment, five performance metrics was utilized on six image datasets.

Based on shape, text and color features, image retrieval process was conducted by Alsmadi (Alsmadi et al., 2020). For this process, they used canny edge histogram and discrete wavelet transform for color feature extraction. Afterwards, text feature extraction was performed by GLCM. Moreover, canny edge method and neutrosophic clustering algorithm was implemented to extract shape features. Finally, Genetic algorithm and simulated annealing technique was implemented to obtain the related pictures based on the similarity. The corel dataset were used for performance evaluation.

Garg and Dhiman (Garg et al., 2021) developed the CBIR framework based on four phases. Initially, the discrete wavelet transformation (DWT) based decomposition was performed for R, G, and B channels. Afterwards, set of functions were used to concatenate these three R, G, and B channels. Then, the feature selection process was conducted by the PSO algorithm which select important features of the image. Finally, the classification process was conducted by three classifiers, which identify the similar pictures from the corel dataset based on the similarity. The effectiveness of this technique was evaluated based on four performance metrics.

A Bi-layer Content Based Image Retrieval (BiCBIR) system was suggested by Singh and Batra (Singh et al., 2020). This BiCBIR system includes two phases. In the first phase, an image's important features like shape, texture, and color are extracted. In the second phase, the classifier contains two layers. Using the texture and shape attributes of the input image, all of the pictures in the database were compared in the first layer. Based on this, the similar images were retrieved and sent to the second layer. In that layer, the images which are similar to the color and shape features are retrieved as the output by the classifier.

Advanced Local Octa-Directional Pattern (ALODP) based CBIR system was suggested by Qasim et al. (Qasim et al., 2022). In this system, principal component analysis (PCA) based feature descriptor was implemented. The intensity of neighbourhood pixels were utilized by the PCA technique in all eight directions. The directional and magnitude patterns are quantized into a 40-bin histogram as a result of the local octa-directional pattern. The magnitude and directional histograms were concatenated to generate the joint histogram. The authors utilized Manhattan distance to check the similarity among images. Moreover, the PCA technique was utilized to decrease the dimensionality and maintain the cost of the computation. The Multi-PIE face dataset was utilized to evaluate the suggested technique based on recall, accuracy and precision performance metrics.

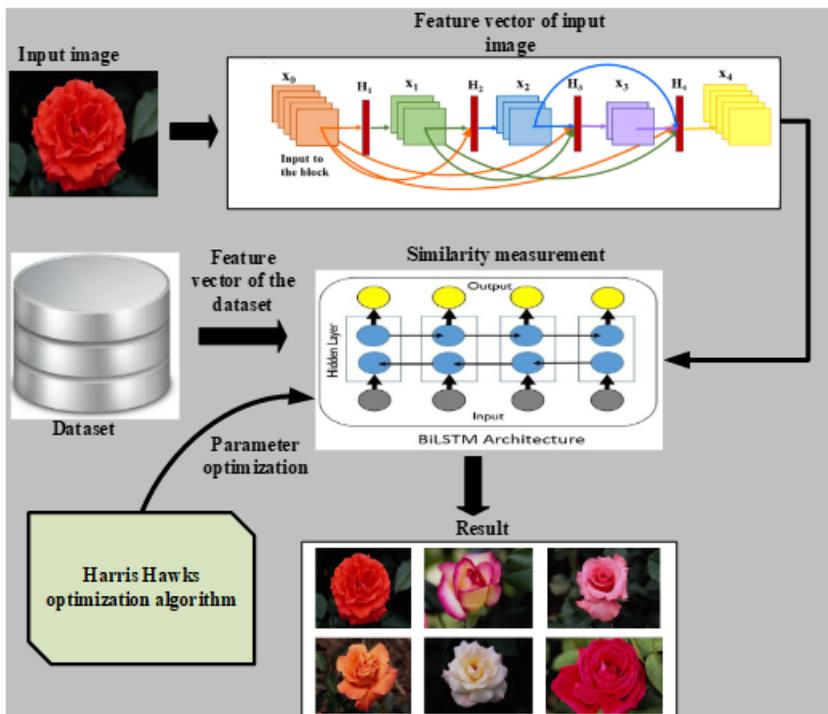
Local feature weighting algorithm was implemented by Ghodrattnama and Abrishami (Ghodrattnama et al., 2021) for CBIR system. This approach groups the data first, then the features are trained to have suitable weights in each cluster and the classification error rate is computed based on these features. In this approach, the supervised c-means technique was implemented to cluster the input images and local feature weighting method were used to weight the image features in each cluster. The cluster that corresponds to the query is located first in the testing phase. In each cluster, the weights of the features were applied to obtain the most similar images. Based on the weights of the features assigned to that cluster, the most similar photos were retrieved.

Proposed Methodology

The important objective of this paper is to propose an effective technique based on hybrid Densenet121-BiLSTM for efficient content based image retrieval system. The proposed framework revealed in Fig. 1 contains of two main sections. They are feature extraction and classification.

Initially the input image is fed into the Densenet 121 for feature extraction. Then these feature vectors are given as an input to the hybrid BiLSTM system. This network is already trained using the Corel dataset. The pre-trained BiLSTM unit will compute the difference among the query image’s features and the complete data set’s features. Based on this difference, the similarity index is created for all images and the images are retrieved which have closer similarity values with respect to query image.

Figure 1: Proposed architecture



Feature Extraction Using Densenet-121

Densely connected convolutional network (DenseNet-121) includes the benefits of both highway networks and Resnet. With the use of this the vanishing gradient problem is reduced in DenseNet. The goal of this network is to ensure maximum information flow among network layers. In a feed-forward method, every layer are directly linked to every other layer. This network includes totally 121 layers with three transition layers and four dense blocks (3-transition, 117-conv, 1-softmax)

DenseBlock is the primary component of DenseNet-121. Convolution Layers make up the DenseBlocks. Generally, CNN architectures are hierarchical. Hence the input of l^{th} layer are the feature maps of $(l - 1)^{th}$ layer. On the other hand, densenet concatenates the feature-maps of all preceding layers and uses them as input for any specific layer. Additionally, it uses its own feature maps as inputs for all following layers.

$$X_l = H_l([X_0, X_1, \dots, X_{l-1}]) \quad (1)$$

Here, the concatenation of the feature maps formed in layer 0, 1, . . . , l – 1 is referred to as $(x_0, x_1, \dots, x_{l-1})$. The function $H(\cdot)$ contains three important operations. They are convolution (Conv), rectified linear unit (ReLU) and batch normalization (BN).

In densenet, the layers that reside among dense blocks are known as transition layers which are responsible for down-sampling using 2×2 average pooling, 1×1 convolution, and batch normalisation. The densenet used ReLU activation function to increase the nonlinearity which is expressed in equation 2.

$$f_l = \text{ReLU}(x) = \begin{cases} x & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2)$$

The features are learned in the convolution layers by feature extraction from the result of preceding layer. A Filter, also known as a convolution kernel, is shared by the extracted features and comprises a set of weights. To boost nonlinearity, the ReLU activation function is used to process the all local weight values and the convolution process is described as follows:

$$z^l = W^l \cdot f_l(z^{(l-1)}) + b^l \quad (3)$$

Here, the bias and weights are denoted by b^l and W^l from $(l-1)^{\text{th}}$ to the l^{th} correspondingly, the activation function is denoted by $f_l(\cdot)$ and the l^{th} layer neuron status is represented by z^l

The pooling layer decreases the dimension of each feature map while keeping the most important data. In this study, global average pooling and max pooling are applied. The convolved features maps are down sampled in the max pool layer by the max pooling operation across 2×2 pixels with 2 strides while keeping the same depth. A global average pooling layer is added after the last dense block, providing all feature map's spatial average at the last convolutional layer. The flatten layer is then used to reorganise the output feature map from Global Average Pooling into a one-dimensional vector, which can then be fed into fully linked layers. The fully connected layer is used to match the learning features' dimensions to the number of categories.

Classification and Similarity Measurement

Bi-directional Long Short Term Memory (BiLSTMs) models are differ from typical LSTMs in that they capture data from both the past and the future. This network can be used in applications where the prediction is dependent on the entire input sequence. This network includes two LSTMs. Here, one LSTM is used to take the input data from end to first layer and another is used to take the input from first to last layer in the series.

BiLSTMs learns the input data in both forward and backwards manners, concatenating the two interpretations depending on long-term data dependencies. BiLSTMs are common deep-learning models with the goal of estimating a function associated with a specific pattern. As a result, a class $y = f * (x, \theta)$ included with the input x and the parameter's (θ) value is acquired by the network, leading to the best function approximation. The features extracted by the Densenet is given as the input to the BiLSTM through the input layer which are further processed by the forward and backward layers.

The positive input sequence is used to generate the output sequence \vec{p}_t from time T-n to T-1 by the forward layer LSTM. Moreover, the input sequence is reversed from T-1 to T-n to generate the backward output \vec{p}_t . Both the process are iteratively performed. The reLU activation function σ_r is

used to process LSTM cell state's backward and forward outcomes. On input sample, the threshold operation is performed by the ReLU function, as illustrated in equation (4), here, the values less than zero are treated as zero.

$$\sigma_r(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (4)$$

Finally, as stated in equation (5), the output layer consists of the sum of both forward and backward LSTM cell state outputs:

$$y_t = \sigma_r(\vec{p}_t + \tilde{p}_t) \quad (5)$$

The final classification output in vector format is represented by $Y_T = \{y_{T-n}, \dots, y_{T-1}\}$. The biLSTM structure is a memory cell that stores gate modules which is used to update the new values in hidden state's parameters. In LSTM cell, the bias vectors are represented by b_o , b_c , b_i , and b_f and the hidden layer weights are represented by W_o , W_c , W_i , and W_f . The following equation shows the $\sigma(t)$ sigmoid function based σ_g gate activation function.

$$\sigma_g = \sigma(t) = \frac{1}{1 + e^{-t}} \quad (6)$$

The forget gate, shown in equation (7), is used to ignore the previous state.

$$f_t = \sigma_g(w_f \cdot [p_{t-1}, x_t] + b_f) \quad (7)$$

In LSTM cell, the transfer of new information is regulated by the input gate i_t , which is expressed in the following equation.

$$i_t = \sigma_g(w_i \cdot [p_{t-1}, x_t] + b_i) \quad (8)$$

The internal recurrence of the cell is controlled by the internal state G_t . It is displayed in equation 9 and 10.

$$\tilde{C}_t = \tanh(W_c \cdot [p_{t-1}, x_t] + b_c) \quad (9)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

As indicated in equations (11) and (12), the information flow to h_t output of LSTM is controlled by the output gate o_t .

$$o_t = \sigma_g(W_o \cdot [p_{t-1}, x_t] + b_o) \quad (11)$$

$$p_t = o_t + \tanh(C_t) \quad (12)$$

Here, p_t is the final output. Three basic BiLSTM parameters, including the Dropout value, the learning rate, and the number of hidden neurons are considered to increase the performance of BiLSTM which are optimized by the effective optimization algorithm explained in the below section.

Parameter Optimization of BiLSTM

In this section, the parameter optimization based on Harris hawks optimization (HHO) is explained. Harris hawks commonly use unexpected attacks to catch the prey. In addition, the Hawks can employ a variety of pursuit methods in response to the changing qualities of the surroundings and prey escape tendencies. In HHO, exploitation, exploration and the change of these two situations will be used to define the algorithm.

Exploration

Harris hawks use their vision to search and locate prey, but finding prey can be complicated. As a result, they will monitor and catch prey using one of two techniques, with Harris hawks having an equal chance of using both. Harris hawks use the first method to develop a unique strategy based on others and random locations. In second strategy, the mean of each agent and current ideal position are used to generate a new result which is expressed in the following formulas

Here, the current iteration is $H(t)$ and in the next iteration the hawk's location is $H(t+1)$. The random individual $H_{rand}(t)$ is chosen from the population and prey's location is represented by $H_{rabbit}(t)$. It is also known as the current optimal individual. The variables $r1, r2, r3, r4$, and q between 0 and 1 make up the upper (UB) and lower (LB) boundary variables. H_m is a representation of the average hawk location in the present population, which may be written as Eq. (13):

$$H_m(t) = \frac{1}{N} \sum_{i=1}^N H_i(t) \quad (13)$$

In the above equation, at t^{th} iteration, the i^{th} Hawks location is $H_i(t)$ and the total number of hawks is denoted by N .

Exploration to Exploitation Transformation

The prey's energy is the important parameter in HHO during transition of exploration to exploitation and its expression is as follows:

$$E = 2E_o \left(1 - \frac{t}{mT} \right) \quad (14)$$

Here, the value of E_o is varied from -1 to 1 for every iteration, the current number of iteration is denoted by t , the maximum number of iteration is denoted by mT and prey's escaping energy is denoted by E . If $|E| \geq 1$, HHO shifts into exploration mode and hawks begin hunting in new areas; if $|E| < 1$, the HHO enter into the exploitation mode and hawks circle the current solution.

Exploitation

In exploitation mode, the prey is captured by the hawks based on the surprise attack. Even though, the prey frequently attempted to leave. Hence the Harris hawks employ a variety of tactics to catch it. Assume that the escape probability before the attack is r . If $r < 0.5$ means the target is safely runaway

from the catch otherwise unsuccessfully ($r < 0.5$). Whatever the situation, the prey is always encircled by the Hawks, whether softly or forcefully, depending on the circumstances. To determine whether the hawks are under attack or not, the E is very important. When $|E|$ is less than 0.5, it besieges lightly; otherwise, it can be a difficult one.

Soft Besiege

The prey try to escape the hawks' encirclement by unpredictable movements when the r and $|E|$ are both > 0.5 , but at the end it is failed. Harris hawks softly attack and drain the prey's energy in this way. The hawks then launch a surprise attack. Eq. (15) provides a mathematical explanation for this phenomenon.

$$H(t + 1) = \Delta H(t) - E \left| JH_{rand}(t) - H(t) \right| \quad (15)$$

$$\Delta H(t) = H_{rabbit}(t) - H(t) \quad (16)$$

$$J = 2(1 - r_5) \quad (17)$$

Here, r_5 is the random number between 0 and 1, The position vector's variation among the hawk and prey is denoted by $\Delta H(t)$ which is expressed in Eq.(16). When the prey try to escape, the force of the random jump is denoted by J which is expressed in Eq.(17).

Hard Besiege

The prey's ability to escape is poor when $r < 0.5$ and $|E| < 0.5$. Even though, the prey is hardly encircled by the hawks and the raid is finally completed. In this case, Eq. (18) is used to update the locations of the present individuals.

$$H(t + 1) = H_{rabbit}(t) - E \left| \Delta H(t) \right| \quad (18)$$

Soft Besiege With Progressive Rapid Dives

The target is safely left with plenty of energy if $r < 0.5$ and $|E| > 0.5$, and the hawks maintain a light siege before hitting. To replicate prey fleeing activity through zigzag deception and unpredictable hawk diving, Levy-flight is incorporated with HHO. Earlier research has shown that under non-destructive foraging conditions, the best search technique is Levy-flight for foragers which is also observed in other animals such as shark and monkey.

According to Eq. (19), the hawks pick the next plan of action:

$$C = Hrabbit(t) - E \left| JHrabbit(t) - H(t) \right| \quad (19)$$

The hawks then conduct a Levy-flight based dive which is shown in Eq. (20). Also, the probable outcome to the preceding one is compared in order to choose the optimal option. It is expressed in Eq.(21).

$$K = C + S \times LF(D) \quad (20)$$

$$H(t+1) = \begin{cases} C & \text{if } F(C) < F(H(t)) \\ K & \text{if } F(K) < F(H(t)) \end{cases} \quad (21)$$

Here, the randomly generated vector with $1 \times D$ size is denoted by S , the dimension of the problem denoted by D and the formulation of Levy-flight is expressed in the following equation which is denoted by LF .

$$LF(x) = \frac{\mu \times \sigma}{|v|^{\frac{1}{\beta}}}, \quad \sigma = \begin{pmatrix} \Gamma(1 + \beta) \times \sin\left(\frac{\pi\beta}{2}\right) \\ \Gamma\left(\frac{1 + \beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)} \end{pmatrix} \quad (22)$$

Here, the constant coefficient β have the value is 1.5 and the v and μ are normally distributed random numbers between 0 and 1.

Hard Besiege With Progressive Rapid Dives

The hawks created the tight encirclement around the prey before striking if $r < 0.5$ and $|E| < 0.5$. The prey's stamina under this circumstance is insufficient to enable escape. The scenario for prey is the same as in a gentle besiege, however, the Hawks are continuously working to reduce the average distance between the prey and themselves. As a result, the subsequent mathematical equations are employed to describe the behaviour of hawk:

$$H(t+1) = \begin{cases} C' & \text{if } F(C') < F(H(t)) \\ Z' & \text{if } F(K') < F(H(t)) \end{cases} \quad (23)$$

Here, Z' and C' is expressed in the following equation:

$$C' = H_{rabbit}(t) - E \left| JH_{rabbit}(t) - H_m(t) \right| \quad (24)$$

$$K' = C' + S \times LF(D) \quad (25)$$

Furthermore, the objective function in Eqs. (21) and (23) is denoted by F and the fitness of vector is denoted $F(*)$ which means the fitness of $C\phi$ is denoted by $F(C\phi)$. Finally, the obtained optimal values are assigned to the parameters of BiLSTM. Therefore, the learning rate is 0.0005, dropout rate is 0.2, and the number of hidden neurons are 100.

Simulation Results

The findings of various tests are designed in this section to analyse the proposed technique's performance and it is reported. Moreover, the explanation of the findings are included in this section. The proposed technique's performance across the Corel dataset is evaluated using F1-score, recall and precision measures. To demonstrate the usefulness of the method for picture retrieval, its performance is compared with the existing state-of-the-art methods. The model is trained and tested on a Windows 10 computer with 16 GB of RAM, and an Intel Ci7 64-bit processor. All the simulations are performed on MATLAB 2018a.

Dataset Description

The investigations in this research used the dataset named Corel, which has 1000 different images, each of which is 256 * 384 or 384 * 256 pixels in size. As a result, the results are reported using ten semantic sets, each including one hundred photos. Flowers, Beach, Horses, Elephants, Food, Buses, Buildings, Dinosaurs, Africa, and Mountains are among the semantic categories in the Corel dataset.

Performance Metrics

F-measure, recall, and precision are the performance parameters used to assess the proposed system's efficiency. The proportion of the amount of recovered relevant pictures (Nr) to the total amount of photos received (Rt) is called precision. The proportion of the amount of recovered relevant photos (Nr) to the full amount of similar pictures (Nt) in the dataset is known as recall. The harmonic mean of recall and precision is known as F-measure. To demonstrate good retrieval performance, recall and precision should be high.

$$Precision = \frac{Nr}{Rt} = \frac{TrPs}{TrPs + FsPs} \quad (26)$$

$$Recall = \frac{Nr}{Nt} = \frac{TrPs}{TrPs + FsNe} \quad (27)$$

$$f - score = 2[(precision * recall) / (precision + recall)] \quad (28)$$

RESULTS AND DISCUSSION

The findings of the presented CBIR framework on Corel data are shown in this section. The F1-score, recall and precision of the proposed technique is computed. Table 1 shows top 10 image's retrieval outcomes of user's query image, as well as the proposed technique's performance on the Corel dataset. The graphical representation of proposed technique is shown in figure 2 and Figure 3 show the findings of relevant photos found by using the elephant image as the query image.

From Table 1 and figure 2, it is observed that the suggested CBIR technique performs better in terms of retrieval. On the corel dataset, the proposed technique attained an average precision of 0.956, an average recall of 0.9306 and average F-score of 0.9216. Particularly, the Elephant and Dinosaur semantic classes had the best f-score, recall and precision values, but the Beach and Africa semantic classes had the weakest f-score, recall and precision values of all 10 classes. The better outcomes on the Dinosaurs and Elephants semantic classes are achieved that both of these categories have unique characteristics and a higher level of differences from other semantic classes, whereas the Africa and Beach semantic classes have similar characteristics and a lower degree of differences from other categories.

Tables 2 and 3 show a category-by-category comparative assessment acquired by the proposed CBIR approach to the outcomes produced by various existing techniques in terms of f-score, recall, and precision scores for each category. Moreover, it is identified from the table that the suggested CBIR approach outperformed previous approaches and demonstrated excellent image retrieval results in the form of f1-score, recall and precision values in a variety of Corel image dataset sets, as well as having the best overall performance results.

Table 1: Results of proposed method for different classes

Semantic class	Precision	Recall	F1-score
Africa	0.983	0.84	0.850
Beach	0.823	0.89	0.825
buildings	0.800	0.95	0.915
buses	1.000	0.965	0.925
dinosaurs	1.000	0.945	0.955
Elephants	1.000	0.967	0.946
Flower	1.000	0.932	0.965
Horse	1.000	0.943	0.964
Mountain	0.954	0.941	0.951
Food	1.000	0.933	0.920
Average	0.956	0.9306	0.9216

Figure 2: Graphical representation of proposed method.

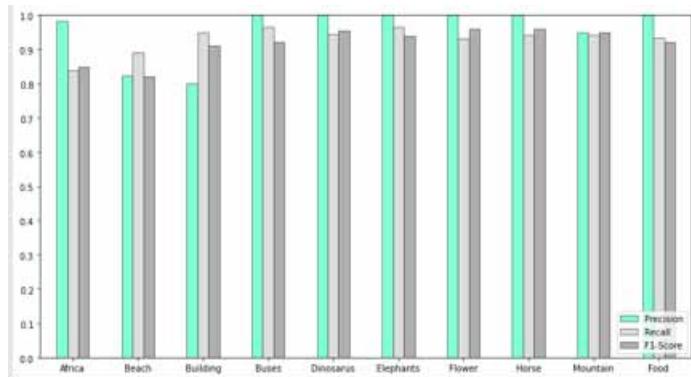


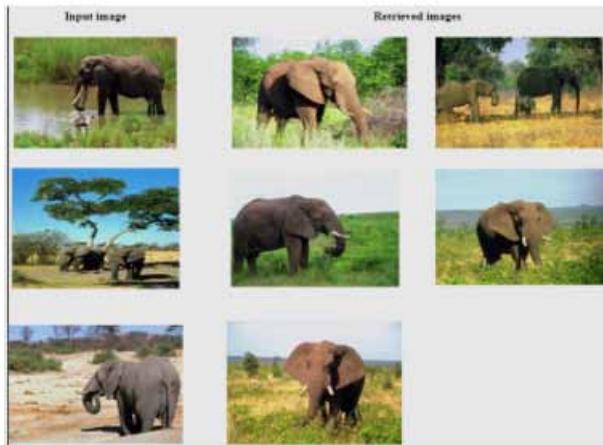
Table 2: Analysis of proposed technique based on precision

Semantic class	SVM (24)	Kmeans(Joseph et al., 2021)	BiCBIR(Singh et al., 2020)	Tetrolet transform+Edge Histogram (Pradhan et al., 2018)	Proposed
Africa	0.812	0.72	0.950	0.950	0.983
Beach	0.9656	0.76	0.750	0.600	0.823
buildings	0.782	0.55	0.850	0.550	0.800
buses	0.833	0.79	1.000	1.000	1.000
dinosaurs	0.822	1	1.000	1.000	1.000
Elephants	0.775	0.7	0.900	0.900	1.000
Flower	0.863	0.87	1.000	1.000	1.000
Horse	0.863	0.7	1.000	1.000	1.000
Mountain	0.902	0.58	0.800	0.750	0.954
Food	0.861	0.56	0.950	1.000	1.000
Average	0.847	0.723	0.920	0.875	0.956

Table 3: Analysis of proposed technique based on Recall and F1-score

Semantic class	SVM (24)		Kmeans (28)		BiCBIR (29)		Tetrolet transform+Edge Histogram (30)		Proposed	
	Recall	F1-score	Recall	F1-score	Recall	F1-score	Recall	F1-score	Recall	F1-score
Africa	0.82	0.815	0.144	0.24	0.190	0.317	0.190	0.317	0.84	0.850
Beach	0.87	0.910	0.152	0.253333	0.150	0.250	0.120	0.200	0.89	0.825
buildings	0.90	0.837	0.11	0.183333	0.170	0.283	0.110	0.183	0.95	0.915
buses	0.80	0.816	0.158	0.263333	0.200	0.333	0.200	0.333	0.965	0.925
dinosaurs	0.79	0.806	0.2	0.333333	0.200	0.333	0.200	0.333	0.945	0.955
Elephants	0.83	0.801	0.14	0.333333	0.180	0.300	0.180	0.300	0.967	0.946
Flower	0.82	0.841	0.174	0.29	0.200	0.333	0.200	0.333	0.932	0.965
Horse	0.82	0.841	0.14	0.233333	0.200	0.333	0.200	0.333	0.943	0.964
Mountain	0.93	0.916	0.116	0.193333	0.160	0.267	0.150	0.250	0.941	0.951
Food	0.87	0.865	0.112	0.186667	0.190	0.317	0.200	0.333	0.933	0.920
Average	0.845	0.8448	0.1446	0.241	0.184	0.307	0.175	0.292	0.9306	0.9216

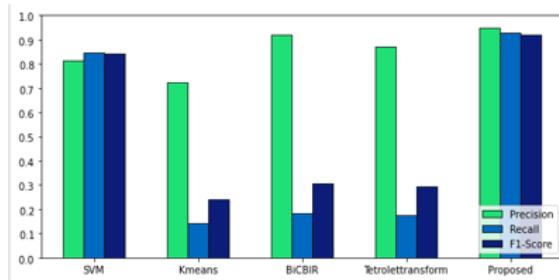
Figure 3: Sample results of query image (Elephant)



Especially, the proposed technique achieves the average F1-score, recall, and precision are 0.91, 0.218, and 0.91 respectively, whereas BiCBIR comes in second with 0.920, 0.184, and 0.307 for average F1-score, recall, and precision respectively. In Fig. 4, the line graph comparison of the suggested and compared approaches are included based on average f-score, recall, and precision.

From the above mentioned experiments, it is clearly seen that the proposed technique outperforms all other strategies, particularly on the eight coefficients where the proposed method obtains average f1-score, recall and precision values of more than 40%. Finally, it is concluded that the proposed method may be utilised to reliably recover photos from the dataset with a wide range of images.

Figure 4: Comparison of proposed method with existing techniques



CONCLUSION

This work proposed the image retrieval system based on deep learning which includes densenet 121 and an optimized BiLSTM to efficiently search databases and retrieve photos that are comparable to the query image. The Corel dataset was used to conduct the experiments and it demonstrate that the proposed approach based similarity measure has great capacity in distinguishing unique characteristics. In different categories of Corel image dataset, the suggested CBIR framework outperformed previous techniques and demonstrated excellent retrieval picture outcomes in the form of precision, recall, and f1-score values, as well as having the highest overall performance result. In future, this approach applied for the medical image retrieval based on their properties.

CONFLICT OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

AVAILABILITY OF DATA AND MATERIAL

Not applicable

CODE AVAILABILITY

Not applicable

AUTHORS' CONTRIBUTIONS

The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

ETHICS APPROVAL

This material is the authors' own original work, which has not been previously published elsewhere. The paper reflects the authors' own research and analysis in a truthful and complete manner.

ACKNOWLEDGEMENTS

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

REFERENCES

- Alshehri, M. (2020). A content-based image retrieval method using neural network-based prediction technique. *Arabian Journal for Science and Engineering*, 45(4), 2957–2973.
- Alsmadi, M. K. (2020). Content-based image retrieval using color, shape and texture descriptors and features. *Arabian Journal for Science and Engineering*, 45(4), 3317–3330.
- Ashraf, R., Ahmed, M., Ahmad, U., Habib, M. A., Jabbar, S., & Naseer, K. (2020). MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features. *Multimedia Tools and Applications*, 79(13), 8553–8579.
- Bibi, R., Mehmood, Z., Yousaf, R. M., Saba, T., Sardaraz, M., & Rehman, A. (2020). Query-by-visual-search: Multimodal framework for content-based image retrieval. *Journal of Ambient Intelligence and Humanized Computing*, 11(11), 5629–5648.
- Chen, D., Chen, Y., Ma, J., Cheng, C., Xi, X., Zhu, R., & Cui, Z. (2021). An ensemble deep neural network for footprint image retrieval based on transfer learning. *Journal of Sensors*.
- Desai, P., Pujari, J., & Sujatha, C. (2021). Impact of multi-feature extraction on image retrieval and classification using machine learning technique. *SN Computer Science*, 2(3), 1–9. doi:10.1007/s42979-021-00532-9
- Desai, P., Pujari, J., Sujatha, C., Kamble, A., & Kamblia, A. (2021). Hybrid Approach for Content-Based Image Retrieval using VGG16 Layered Architecture and SVM: An Application of Deep Learning. *SN Computer Science*, 2(3), 1–9. doi:10.1007/s42979-021-00529-4
- Garg, M., & Dhiman, G. (2021). A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants. *Neural Computing & Applications*, 33(4), 1311–1328.
- Ghodratnama, S., & AbrishamiMoghaddam, H. (2021). Content-based image retrieval using feature weighting and C-means clustering in a multi-label classification framework. *Pattern Analysis & Applications*, 24(1), 1–10.
- Hidayat, B. M. H., & Putra, R. E. (2019). Penerapan CNN dengan Filter Gabor sebagai feature extractor untuk Content-Based Image Retrieval. *Journal of Information and Computational Science*, 1(01).
- Hu, H., Zheng, W., Zhang, X., Zhang, X., Liu, J., Hu, W., Duan, H., & Si, J. (2021). Content-based gastric image retrieval using convolutional neural networks. *International Journal of Imaging Systems and Technology*, 31(1), 439–449. doi:10.1002/ima.22470
- Joseph, A., Rex, E. S., Christopher, S., & Jose, J. (2021). Content-based image retrieval using hybrid k-means moth flame optimization algorithm. *Arabian Journal of Geosciences*, 14(8), 1–14.
- Keisham, N., & Neelima, A. (2022). Efficient content-based image retrieval using deep search and rescue algorithm. *Soft Computing*, 26(4), 1–20. doi:10.1007/s00500-021-06660-x
- Khalid, M. J., Irfan, M., Ali, T., Gull, M., Draz, U., Glowacz, A., & Hussain, S. et al. (2020). Integration of discrete wavelet transform, DBSCAN, and classifiers for efficient content based image retrieval. *Electronics (Basel)*, 9(11), 1886.
- Khan, U. A., Javed, A., & Ashraf, R. (2021). An effective hybrid framework for content based image retrieval (CBIR). *Multimedia Tools and Applications*, 80(17), 26911–26937.
- Kumar, R. (2022). A hybrid feature extraction technique for content based medical image retrieval using segmentation and clustering techniques. *Multimedia Tools and Applications*, 1–34.
- Öztürk, Ş. (2020). Stacked auto-encoder based tagging with deep features for content-based medical image retrieval. *Expert Systems with Applications*, 161, 113693.
- Pathak, D., & Raju, U. S. N. (2021). Content-based image retrieval using feature-fusion of GroupNormalized-Inception-Darknet-53 features and handcraft features. *Optik (Stuttgart)*, 246, 167754.
- Pradhan, J., Ajad, A., Pal, A. K., & Banka, H. (2020). Multi-level colored directional motif histograms for content-based image retrieval. *The Visual Computer*, 36(9), 1847–1868.

- Pradhan, J., Kumar, S., Pal, A. K., & Banka, H. (2018). A hierarchical CBIR framework using adaptive tetrolet transform and novel histograms from color and shape features. *Digital Signal Processing*, 82, 258–281.
- Putzu, L., Piras, L., & Giacinto, G. (2020). Convolutional neural networks for relevance feedback in content based image retrieval. *Multimedia Tools and Applications*, 79(37), 26995–27021. doi:10.1007/s11042-020-09292-9
- Qasim, M., Mahmood, D., Bibi, A., Masud, M., Ahmed, G., Khan, S., & Hussain, S. J. et al. (2022). PCA-Based Advanced Local Octa-Directional Pattern (ALODP-PCA): A Texture Feature Descriptor for Image Retrieval. *Electronics (Basel)*, 11(2), 202.
- Sezavar, A., Farsi, H., & Mohamadzadeh, S. (2019). Content-based image retrieval by combining convolutional neural networks and sparse representation. *Multimedia Tools and Applications*, 78(15), 20895–20912. doi:10.1007/s11042-019-7321-1
- Sharif, U., Mehmood, Z., Mahmood, T., Javid, M. A., Rehman, A., & Saba, T. (2019). Scene analysis and search using local features and support vector machine for effective content-based image retrieval. *Artificial Intelligence Review*, 52(2), 901–925. doi:10.1007/s10462-018-9636-0
- Singh, S., & Batra, S. (2020). An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications*, 79(25), 17731–17759.
- Singh, S., & Batra, S. (2020). An efficient bi-layer content based image retrieval system. *Multimedia Tools and Applications*, 79(25), 17731–17759.
- Srivastava, P., & Khare, A. (2019). Content-based image retrieval using local ternary wavelet gradient pattern. *Multimedia Tools and Applications*, 78(24), 34297–34322.
- Sundararajan, S. K., Sankaragomathi, B., & Priya, D. S. (2019). Deep belief CNN feature representation based content based image retrieval for medical images. *Journal of Medical Systems*, 43(6), 1–9. doi:10.1007/s10916-019-1305-6 PMID:31069547
- Zhang, K., Qi, S., Cai, J., Zhao, D., Yu, T., Yue, Y., Yao, Y., & Qian, W. (2022). Content-based image retrieval with a Convolutional Siamese Neural Network: Distinguishing lung cancer and tuberculosis in CT images. *Computers in Biology and Medicine*, 140, 105096. doi:10.1016/j.combiomed.2021.105096 PMID:34872010
- Zhong, A., Li, X., Wu, D., Ren, H., Kim, K., Kim, Y., Buch, V., Neumark, N., Bizzo, B., Tak, W. Y., Park, S. Y., Lee, Y. R., Kang, M. K., Park, J. G., Kim, B. S., Chung, W. J., Guo, N., Dayan, I., Kalra, M. K., & Li, Q. (2021). Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19. *Medical Image Analysis*, 70, 101993. doi:10.1016/j.media.2021.101993 PMID:33711739

K. Sanjeevaiah is an Assistant Professor in Department of Computer Science and Engineering in Hyderabad Institute of Technology and Management, Basuragadi Village, Hyderabad, Telangana, India.

Tatireddy Subba Reddy is an Associate Professor in Department of Computer Science and Engineering, B V Raju Institute of Technology Narsapur, Telangana State, India.

Sajja Karthik is an Assistant Professor in Computer Science and Engineering, R.V.R. & J.C. College of Engineering, Chowdavaram Guntur, Andhra Pradesh, India.

K. Mahesh Kumar is an Associate Professor in Department of Computer Science and Engineering, Mahatma Gandhi Institute of Technology, Hyderabad, India.

D. Vivek is an Associate professor in Computer Science and Engineering, B. V. Raju Institute of Technology, Narsapur, Telangana, India.