The Scientific Temper (2024) Vol. 15 (spl): 415-424
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2024.15.spl.48
https://scientifictemper.com/

RESEARCH ARTICLE

Retrieval-Based Inception V3-Net Algorithm and Invariant Data Classification using Enhanced Deep Belief Networks for Content-Based Image Retrieval

Shaik Abdulla P.*, Abdul Razak T.

Abstract

In the present scenario, Content-Based Image Retrieval (CBIR) performs a constantly changing function that makes use gain knowledge from images. Moreover, it is also the dynamic sector of research and was recently rewarded due to the drastic increase in the performance of digital images. To retrieve images from the massive dataset, experts utilize Content Based Image Retrieval. This approach automatically indexes and retrieves images depending upon the contents of the image, and the developing techniques for mining images are based on the CBIR systems. Based on the visual characteristics of the input image, object pattern, texture, color, shape, layout, and position classifications are applied, and indexing is carried out. When issues arise during feature extraction, deep learning approaches help to resolve them. A method called RIV3-NET, which stands for Retrieval-Based Inception V3, was used to classify the features. Classifying image invariant data using Enhanced Deep Belief Networks (EDBN) is necessary to decrease noise and improve displacement with smoothness. The simulation outcomes demonstrate the improved picture retrieval and parametric analysis.

Keywords: Content-based image retrieval, Deep learning, Retrieval inception V3-NET algorithm, Enhanced deep belief networks.

Introduction

To efficiently carry out the function of the classifier, the system that obtains the image sorts the database images into two groups: relevant and irrelevant. Several supervised learning techniques have been employed to deal with the two main challenges of classification. The first challenge is the limitation in labeled or annotated training samples. In general, queries provide labels and not much relevant feedback. As training samples are limited, classification is weaker. The next challenge is related to the dimension of the visual data. Weighting features, making selections, and reducing dimensionality all become more difficult when it's high. Dimensionality reduction strategies that make the most of small training sets are not immune to this problem:

1. Dharani, T., & Laurence Aroquiaraj, I. (2013).

Image mining is a growing field that extracts knowledge from images by browsing, searching, and identifying them from a large digital database. It involves extracting image features, comparing the query image with the database, and displaying similar images for fast matching. Complex images can be difficult to retrieve, so computer-based image retrieval (CBIR) systems use attributes like color, texture, and spatial layout to denote images. Images are also filtered based on content for better indexing and more accurate results, ManickaChezian, R., & Janani, M. (2012).

CBIR systems extract low-level features from images, either from the entire image or specific portions. They are region-based due to user interest in specific regions. Global feature retrieval systems are simple and represent images at the regional level, like human perception. To guarantee image similarity, segmented areas can be used to extract region-based information such as color, shape, texture, and spatial placement, Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008).

PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli, Tamilnadu, India.

***Corresponding Author:** Shaik Abdulla P., PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University, Tiruchirappalli), Tiruchirappalli, Tamilnadu, India., E-Mail:

How to cite this article: Abdulla,. S.P., Razak, A.T. (2024). Retrieval-Based Inception V3-Net Algorithm and Invariant Data Classification using Enhanced Deep Belief Networks for Content-Based Image Retrieval. The Scientific Temper, **15**(spl):415-424.

Doi: 10.58414/SCIENTIFICTEMPER.2024.15.spl.48

Source of support: Nil

Conflict of interest: None.

© The Scientific Temper. 2024

Received: 11/09/2024 **Accepted:** 28/09/2024 **Published:** 16/10/2024

The texture-based approach re-groups image texture features using statistical parameters like contrast, entropy, dissimilarity, auto-correlation, mean, standard deviation, and variance. The image is identified from the database using these values. To estimate textures, the grey level co-occurrence matrix (GLCM) successfully collects second-order statistics from the image.

There are three levels of visual characteristics: primal, logical, and abstract. Features such as color and shape are considered primitive, whereas features such as object identification and importance are considered logical. On the other hand, modern algorithms only employ rudimentary characteristics when visual cues are paired with human explanations. Even systems like Blob World cannot reliably identify objects due to the semantic gap, which is the loss of image information to be represented as features. This gap is problematic when using image retrieval applications but can be smaller with domain knowledge. The sensory gap also occurs, causing a loss between the original structure and digital image representation, Rehman, M., Iqbal, M., Sharif, M., & Raza, M. (2013).

Related Works

The authors propose a new image retrieval method combining TF-IDF and CNN which has been developed for analyzing visual content. The model's performance was evaluated on four datasets, and a hashing algorithm was developed for large-scale datasets. The code used deep learning methods to generate binary representations and extract features. Experiments on histopathology images showed an impressive classification accuracy of 97.94%, demonstrating its reliability in handling large-scale datasets, Kondylidis, N., Tzelepi, M., & Tefas, A. (2018).

The authors propose a pairwise-based deep-ranking hashing (PDRH) algorithm for histopathology image analysis. It extracts features and learns binary representations, preserving inter-class differences for classification and intra-class relevance order for retrieval. The algorithm's effectiveness and efficiency were validated on a large dataset of histopathological skeletal muscle and lung cancer images, demonstrating high classification accuracy and retrieval performance, Shi, X., Sapkota, M., Xing, F., Liu, F., Cui, L., & Yang, L. (2018).

The authors developed semantics-assisted visual hashing (SAVH), an unsupervised model that converts image pixels into mathematical vectors using extracted texture and visual features. Text is extracted using a topic hypergraph, and semantic details are derived. The image's hash code is examined to preserve the correlation between images and semantics, and a hash function is generated. This feature is crucial for real-time applications of CBIR, Zhu, L., Shen, J., Xie, L., & Cheng, Z. (2017).

The author's CNNs are highly effective in computer vision applications, particularly in CBIR approaches. However, they often rely on intermediate convolutional layers to identify local patterns. A new technique called bilinear CNN architecture uses two CNN models in parallel for feature extraction without knowing the image's semantics. This reduces image representation and enhances performance, search time, and storage cost. It can analyze complex images with distinct semantics and provides superior performance when applied to larger databases, Alzu'bi, A. A., & Ramzan, N. (2017).

The authors suggested that the "Supervised Learning of Semantics-Preserving Hash via Deep Convolutional Neural Networks" presents the SSDH approach, which builds binary hash codes from labeled data for effective large-scale picture retrieval. Unifying classification and retrieval inside one learning model and scalable to big datasets, SSDH minimizes an objective function that encompasses classification error and desirable hash code features, Yang, H. F., Lin, K., & Chen, C. S. (2018).

The authors pioneered "ImageNet Classification with Deep Convolutional Neural Networks," a game-changing achievement in computer vision that involved training a CNN to categorize more than one million high-resolution photos into one thousand separate classes. The model surpassed prior methods with top-1 and top-5 error rates of 37.5% and 17.0%, respectively, because to its 60 million parameters and 650,000 neurons. The "dropout" regularization method, optimized GPU implementation, and Rectified Linear Units were important advances, Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012).

The authors presented a strategy for effectively learning face verification's high-level features. These features are derived from multi-scale mid-level features and are constructed on the feature extraction hierarchy of deep ConvNets. This technique learns properties that are both compact and discriminative by representing many identities with few hidden variables. Complementary features derived from other facial areas significantly improve performance. On the LFW dataset, the suggested technique was able to obtain a face verification accuracy of 97.45% with faces that were weakly aligned, Sun, Y., Wang, X., & Tang, X. (2014).

The authors presented a model that can use a dataset of texts and images to create natural language descriptions of image regions using weak labels. The innovative ranking approach achieves state-of-the-art results in image-sentence ranking trials by integrating visual and verbal modalities using a shared multimodal embedding. Furthermore, a design for multimodal recurrent neural networks (RNNs) that can explain visual data is detailed. Results from full-frame and region-level trials show that the RNN model is superior to retrieval baselines, Karpathy, A., & Fei-Fei, L. (2015).

The authors presented a new approach for social picture interpretation called deep collaborative embedding (DCE) that combines collaborative factor analysis with end-to-end

learning. It handles out-of-sample issues, factors multi-correlation matrices, and improves tagging matrices. Some of the uses for the model include expanding tags, improving tag refinement, and retrieving images based on content, Li, Z., Tang, J., & Mei, T. (2018).

The authors propose a multi-view label sharing (MVLS) model was developed to preserve and maintain similarity in visual representation and classification. The model was tested using six and nine views, and its effectiveness was demonstrated when compared to standard methods, demonstrating its effectiveness in visual representation and classification, Zhang, C., Cheng, J., & Tian, Q. (2018).

The authors developed and compared CNNs and local features that have been used in image understanding and object classification, but they face challenges due to precise object classification and limited training data. The MVFL-VC method overcomes this by consistently employing both labeled and unlabelled images. Experiments showed the MVFL-VC method's superiority over other image classification and representation methods, demonstrating its effectiveness across various unlabeled and unseen datasets, Zhang, C., Cheng, J., & Tian, Q. (2019).

The authors propose that people's eyes are very sensitive to little changes in color. The suggested RGB images were transformed into YCbCr color space in order to record finer details. In order to retrieve images from queries, feature vectors were created using the extracted edge features from the Canny edge detector. To decrease the number of computing steps, histogram and Haar wavelet transforms were employed. Results were good when comparing the performance of an Artificial Neural Network (ANN) with that of existing CBIR systems, Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., & Jeon, G. (2018).

The authors examined the feasibility of merging SIFT feature indexing with deep convolutional neural networks (d-CNN) for picture retrieval and developed the method for collaborative index embedding. The authors came up with a method they dubbed the collaborative index embedding technique that continuously updated the index of CNN and SIFT features. This allowed them to implicitly merge CNN with SIFT features while still ensuring the neighborhood structure of the shared image. Applying a CNN-embedded index to online searches after iterative index embedding improves retrieval accuracy by 10% compared to the actual CNN and SIFT index. This technique demonstrated superior performance while retrieving photos, Zhou, W., Li, H., Sun, J., & Tian, Q. (2018).

The authors deployed a CBIR system that employs ML to sift through database photos for feature vectors, classify 60–70% of images in each class, and train a classifier. The online step involves the user inputting a query image, and the classifier predicting the name based on the feature vector calculated using Local Patterns, Wiggers, K. L., Britto, A. S., Heutte, L., Koerich, A. L., & Oliveira, L. E. S. (2018).

The authors assessed the CBIR system by means of Corel databases and three machine learning techniques: SVM, KNN, and CNN. The study sheds light on the efficacy of deep learning, KNN, and CNN algorithms by comparing their accuracy and efficiency in particular picture retrieval tasks, Yenigalla, S. C., Rao, K. S., & Ngangbam, P. S. (2023).

The authors presented a novel CBIR system that utilizes LNP and ML approaches; this system outperformed LBP, LDP, and LTrP in terms of average recall, and when paired with LNP, it improved average accuracy, Alrahhal, M., & Supreethi, K. P. (2019).

The authors developed the CBIR method—which makes use of color, shape, and texture that is rapidly becoming a standard in multimedia systems for automated picture retrieval and speedier searching and it is finding use in areas like surveillance detection, crime prevention, and fingerprint matching, Koyuncu, H., Dixit, M., & Koyuncu, B. (2021).

The authors learned that the expansion of digital image sensors and the prevalence of the internet have heightened the demand for effective search strategies for retrieving images. With an outline of the CBIR architecture, low-level feature extraction methods, machine learning algorithms, similarity metrics, and performance evaluation, this compares current methodology in CBIR and should encourage additional study in the field, Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021).

The authors brought CBIR systems to computer vision, which tackled issues like scalability and semantic gaps. It investigates learning methodologies, ML, DL, and convolutional neural networks in an effort to enhance CBIR performance and suggests solutions including relevance feedback, Qazanfari, H., AlyanNezhadi, M. M., & Nozari Khoshdaregi, Z. (2023).

The authors put forward the primary function of image retrieval systems is to search through large databases for specific images. Images that are visually and semantically comparable to a query image are the primary focus of CBIR approaches. Scientists have come up with a new way to retrieve photos by measuring their independence using histograms, statistical features, and the T-test, Ali, F. (2020).

The author's research in multimedia CBIR systems has recently been boosted by technology improvements, which have increased the complexity of multimedia. These systems try to extract images from enormous databases, but their usefulness is constrained by the sets of features they have. This technique uses a gray-level co-occurrence matrix, a neutrosophic clustering algorithm, a Canny edge detection approach, and RGB color to extract robust features from feature vectors, Alsmadi, M. K. (2020).

The authors indicated that a major obstacle to research in the field of retrieving similar content is the complexity of multimedia. In order to retrieve material from the internet,

Multimedia Indexing Technology is essential. An innovative CBIR method integrates texture and color information to derive vectors of local features. Feature extraction, similarity matching, and performance evaluation are all parts of the research. The suggested method is great at feature extraction, Ashraf, R., Ahmed, M., Ahmad, U., Habib, M. A., Jabbar, S., & Naseer, K. (2020).

Proposed Methodology

The suggested approach to content-based picture retrieval is detailed in this section. Its overall architecture is illustrated in Figure 1 as follows:

Whenever the input has been preprocessed, the training step begins with feature extraction. Feature extraction and preparation of the input query image are performed in the testing phase. After that, invariant data was identified by classifying the trained and test output. It is possible to get the picture by using the invariant data categorization results.

The proposed CBIR technique's implementation architecture is shown in Figure 2. At first, the input image has been preprocessed for resizing, removing the noise, and enhancing its displacement with smoothness. Then the feature is extracted using the RIV3-NET algorithm, and once the trained output and test output are obtained, the classification for invariant data of an image is done using EDBN. After classifying by ranking matrix, the appropriate image has een obtained with higher accuracy. Finally, the classified output has been obtained.

Feature extraction with the retrieval-based inception V3-NET approach

For providing better performance of the deep CNN approach, the depth as well as the width of the network must be increased, which will also increase the network parameters. This can be enhanced even more by using the GooLeNet model, where the Inception structure is introduced. Maintaining the network model sparse and the dense matrix's excellent computational performance, the key goal is to find the best local sparse framework

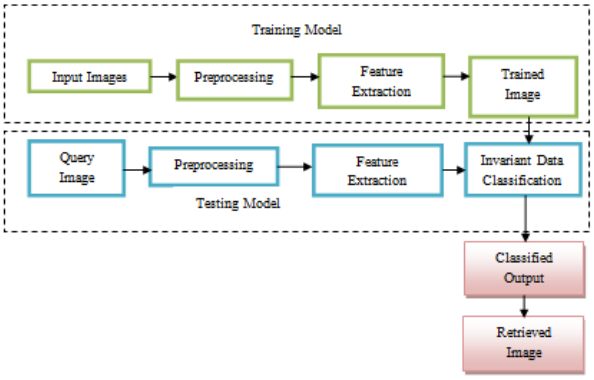


Figure 1: Fundamental Architecture of the Proposed Approach

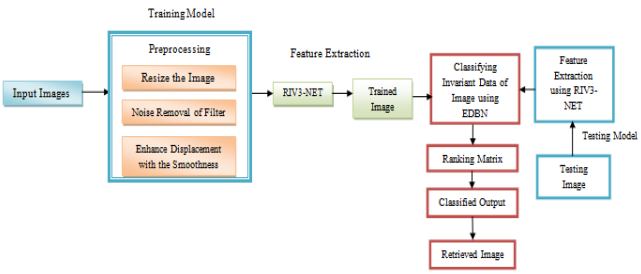


Figure 2: Implementation of Proposed Methodology

with dense components. The central part of the Inception structure contains three Inception modules, whose structure is depicted in Figure 3. These modules' function is to use 1x1 convolutions to lower the dimensionality of the feature map from the prior layer.

The next process is the extraction of features using 1x1, 1xn, and 3x3 convolutional layers; the Filter concat layer, the last layer, comprises LRN (Local Response Normalization) and a Depth concat layer. The Depth concat layer combines the features extracted with convolutional layers. Expanding the Inception structure's network and increasing the ratio of its convolutional layers will lead to an increase in feature channels. While selecting large convolutional kernels like 3x3 and 5x5, computation is greatly increased. Thus, they are mostly small, such as 1x1 and 1xn, which reduces the computation. In addition, the Inception architecture incorporates two additional softmax layers for forward propagation and uses an average pooling layer instead of a fully linked one.

Using a convolution kernel splitting technique to divide big volume integrals into smaller convolutions, Inception v3's network architecture differs from Inception v1 and v2. As an example, 3*3 convolutions are partitioned as 3*1 and 1*3. By using this approach, parameter count is reduced, thereby accelerating the network training speed while spatial features are effectively extracted. Simultaneously, Inception v3 optimizes the network structure of the Inception module with three area grids of various sizes, such as 35*35, 17*17, and 8*8, as illustrated in Figure 4.

An input layer, an output layer, and a number of hidden layers make up a convolutional neural network (CNN).

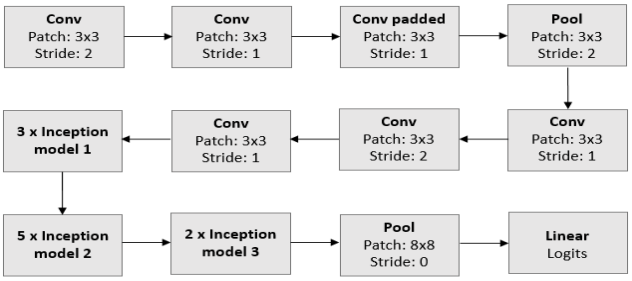


Figure 3: Structure of Inception module

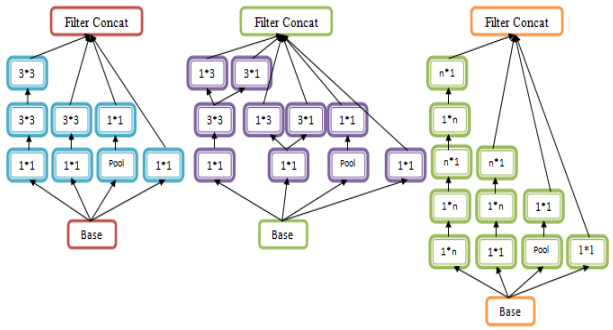


Figure 4: Inception module in RIV3-NET

In general, hidden layers comprise four layers, namely convolutional, pooling, ReLU (normalization), and fully connected (FC) layers. Even a few more layers can be included when the models are too complex. CNN has proved its excellence in several problems related to Computer Vision and Machine Learning. At the abstract level, CNN does train and prediction with the concepts provided for succeeding sections. This model is widely used in recent Machine Learning applications as it produces effective outcomes.

The operation of CNN is based on linear algebra. Data and weights are represented like matrix-vector multiplication. Every layer holds various characters' sets for an image. Consider if an image of a face is provided as an input to CNN with its initial layers. CNN will determine a few primary features like edges, dark and bright spots, shapes, and so on. The next layer recognizes shapes and objects related to the image, like the mouth, nose, and eyes. The succeeding layers identify the objects that are similar to their actual faces. For matching, CNN considers parts instead of the entire image. Hence the entire image is divided into smaller parts for the process of classifying images. CNN represents the extracted features as a 3x3 grid for evaluation. The filtering process organizes the feature with the image patch. Every pixel is multiplied one by one with the respective feature pixel. After completing this process, the average of all these values is obtained. This final feature value is added to the feature patch. This process is continued for every other feature patch. Moreover, every possible match is tried for this filter, which is termed convolution. Figure 5 displays the network architecture of the RIV3-NET using CNN.

The next CNN layer, "max pooling," reduces the size of the image stack. Setting the window size and stride is necessary for image pooling. After noting each window's maximum value, the image is filtered across the window in strides. For every feature map, the dimensionality is reduced using the Max pooling layer; at the same time, the most valuable information is retained. The process of ReLU of CNN, also termed a normalization layer, changes every negative value to 0 within the filtered image. This step is performed on every filtered image, and thus the non-linear properties of the model are increased by the ReLU layer.

The next step of CNN is layer stacking, which involves convolution, pooling, and ReLU layers, as depicted in Figure 6. Consequently, each layer's output is used as an input for the subsequent layer. When layers are repeated, it leads to "deep stacking." In the CNN architecture, the fully connected (FC) layer, termed as classifier, is the final layer. Every value in this layer gets a vote by classifying the image. Often, these FC layers are stacked, where every intermediate layer votes on the phantom "hidden" layer. Consequently, every additional layer helps the network to even understand more complex combinations of features to make better decisions. Using back propagation through a deep neural network, the weights for FC layers and values for the convolution layer are obtained. Back propagation utilizes the error in the result to estimate the adjustments and changes in the network.

The Inception-v3 model is one of the most accurate models for classifying images and achieves a 3.46% "top-5 error rate" when trained on the ImageNet dataset. This model is widely involved in various tasks like detecting objects and other areas through Transfer Learning.

Invariant Data Classification Using Enhanced Deep Belief Networks (EDBN)

One approach is the use of deep learning networks, which use multiple processing layers to abstract data at a high level. Consequently, these methods work for complicated issues that are semi-supervised as well as unsupervised.

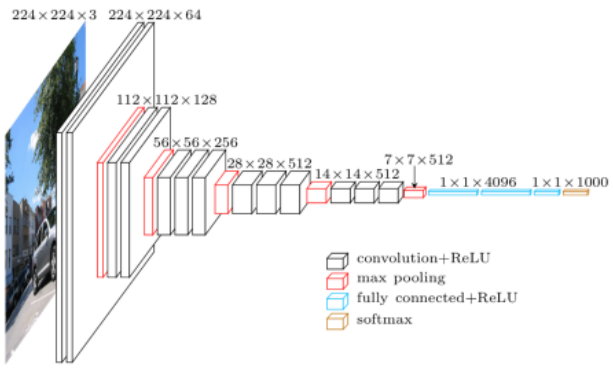


Figure 5: RIV3-NET Network Structure with CNN

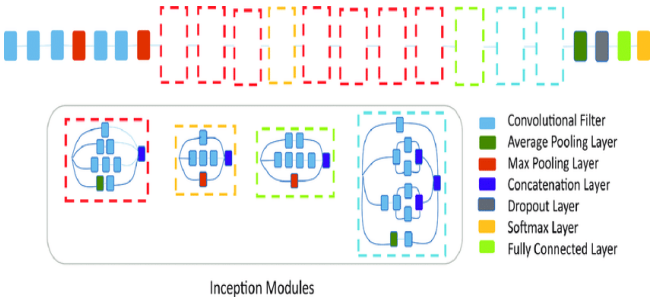


Figure 6: RIV3- NET Layers

The ability to do unsupervised pre-learning is a key feature of deep neural networks, which form the basis of a deep belief network (DBN). DBN is made up of many RBMs, or Restricted Boltzmann Machines. At the outset, DBN employs greedy and unsupervised learning layer-wise, with RBM representing each layer. The inputs are reconstructed by DBN unsupervised training, which does not use target labels. Fig. 7a shows that each RBM receives its output as input from the RBM above it. For RBMs without inter-layer connections, the building blocks are two layers: visible and stochastic hidden units. Both stochastic binary and Gaussian real-valued units are in use. To train RBMs, the Contrastive Divergence (CD) algorithm, namely CD-1, is employed in a three-stage process.

Figure 7b illustrates the supervised DBN training process. In DBN, every layer is a feature generator where the input is converted to a more abstract representation. DBN after unsupervised learning is modified by supervised learning with target labels for classification or regression using gradient descent. Deep neural networks are widely applied in applications like devising predictive feature space for detecting objects from natural images.

Equation 1 estimates the hidden layer values using the posterior probability distribution and the visible units that are provided.

$$p\left(h_j = 1|v,\theta\right) = \sigma\left(a_j + \sum_{i=1}^{V} w_{ij} v_i\right) \qquad Eq.(1)$$

where θ = (w; b; a) It stands for the RBM parameters, which are weights, visible biases, and hidden biases, in that order. $\sigma(x) = \left(1 + e^{-x}\right)^{-1}$ is the number of visible units, and V is the sigmoid function. The values of the visible units are then recreated using the hidden units that have been provided. The posterior probability of the reconstructed values, depending on the type of visible units, will be,

$$p\left(v_j = 1|h,\theta\right) = \sigma\left(b_j + \sum_{j=1}^{H} w_{ij} h_j\right) \qquad Eq.(2)$$

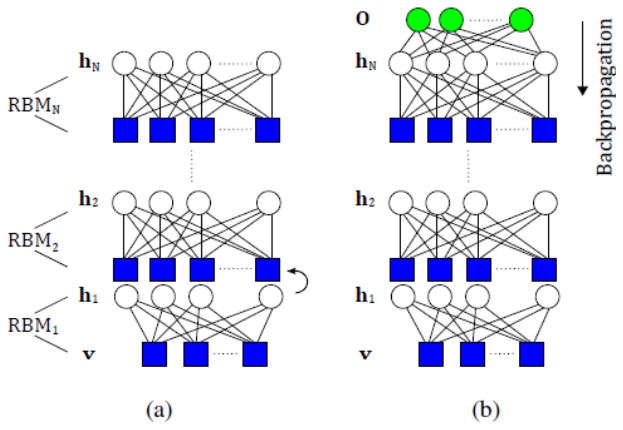


Figure: 7a. Unsupervised and 7b. Supervised DBN training

$$p\left(h_j = 1|v,\theta\right) = \mathcal{N}\left(b_j + \sum_{j=1}^{H} w_{ij} h_j, 1\right) \qquad Eq.(3)$$

where H represents the number of hidden units. $\mathcal{N}\left(\mu,\delta^2\right)$ is a Gaussian with mean μ and variance δ^2. Before using hidden unit likelihoods in equations 2 and 3, they are transformed to binary values. Finally, for the values of reconstructed visible units, the initial step is repeated. Once all the steps are completed, the weights of the network are modified by using equation 4.

$$\triangle w_{ij} \approx -\varepsilon\left(v_i h_{j\,data} - v_i h_{j\,recons}\right) \qquad Eq.(4)$$

where ε and w_{ij} represent the learning rate and weight between a pair of visible units v_i and hidden units h_j respectively. $\cdot_{data}$ and $\cdot_{recons}$ indicates the expectations when the values of the hidden state are derived from the input and reconstructed data, respectively. The process is continued until the algorithm satisfies the required condition, with each repetition referred to as an epoch. In order to expedite the parameter update process, the complete training dataset is divided into smaller subsets known as mini batches. The generative DBN, depicted in Figure 7a, can be transformed into a discriminative model. The process of achieving this transformation involves the addition of a label layer positioned at the top of the network, followed by the implementation of a conventional backpropagation method, as depicted in Figure 7b. Typically, in the case of discriminative DBN, the pre-training stage involves using a greedy RBM-based layer-wise training approach.

Steps of EDBN Algorithm:

Step 1: The initial layer is trained as a Restricted Boltzmann Machine (RBM) which generates unprocessed input. $x = h^{(0)}$ to its exposed layer.

Step 2: The initial layer is utilised to acquire the input representation that is subsequently employed as data in the following layer. There are two commonly used solutions that can serve as the average activations.

$p(h^{(1)} = 1 \mid \mathrm{h}^{(0)})$ or samples of $p(h^{(1)} \mid \mathrm{h}^{(0)})$

Step 3: The second layer is trained using Restricted Boltzmann Machines (RBMs), where the modified data is treated as the training examples.

Step 4: Steps 2 and 3 are iterated for the required number of layers, consistently spreading either samples or mean values in an upward direction.

Step 5: Every parameter of this model is modified related to the proxy for EDBN log- likelihood, or supervised training criterion once additional learning machinery is added up for converting learned representation into supervised ones.

Ranking Matrix:

A ranking matrix f ∈ F reflects the combined value for each of the output layers of the EDBN model, and it approximates the overall value for each instance. $x_i' \in X$. Therefore, it is possible to establish a direct relationship between the output pixel values of the model output and a set of ranking functions. Equation 5 calculates and assigns a score to each item. The comparison function can be established according to diverse parameters, contingent upon the specific context.

$$x'_i > x'_j \Leftrightarrow f(x_i') f(x_j') \qquad Eq.(5)$$

The ranking is calculated by analyzing the values of pairs of pixels (x_i', x_j') The problem at hand is commonly regarded as a learning problem and is derived from the ranking problem. It is extensively employed for classifying pairs of pixel (x_i', x_j') based on their rankings, rather than using the best or worst ranked values. The relationship between (x_i', x_j') is provided by a newly introduced vector (x_i', x_j') in a more established manner.

$$\left(x_i' - x_j', z = \begin{cases} +1 & y_i > y_j \\ -1 & y_j > y_i \end{cases} \right) \qquad Eq.(6)$$

Consequently, two categories are designated to categorize every pair of photos(x_i', x_j'). When the samples are accurately sampled and ranked (+1), two classes are considered positive. The alternative class is assigned a negative label (-1) when the samples are mistakenly classified. Furthermore, x_i' should precede x_i' with the former referring to cases where the transposition is valid. The equation above consists of $x_i' \in$ where each pixel value belongs to the original instances and creates a new instance in the training dataset S'. It generates new labelled vectors based on equation 6.

Results and Discussion

Performance Analysis

This section presents a performance analysis of the proposed RIV3-NET and EDBN. RIV3-NET demonstrates superior performance in image retrieval by providing optimised output with increased efficiency. The performance of the Content-Based Image Retrieval (CBIR) system is evaluated using several metrics, including precision, recall, F1 score, and accuracy. These metrics are defined as follows:

Equation 7 defines precision as the probability of true positives, which refers to correctly identified real positives. The formula for precision is presented as:

$$Precision = \frac{TP}{TP + FP} \quad Eq.\ (7)$$

Recall is defined as the ratio of Real Positives which are correct Predicted Positive given by equation 8

$$Recall = \frac{TP}{TP + FN} \quad Eq.\ (8)$$

F1 score is a metric calculated using equation 9, both precision and recall, which is defined as

$$F1\,Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad Eq.\ (9)$$

The accuracy is calculated using equation 10.

Table 1: Analysis of Image Retrieval

Input Image	*Retrieved Image*			
	CNN	*KNN*	*LBP*	*RIV3-NET-EDBN*
 Ship image				
 Parachute image				
 Signboard image				

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad Eq.(10)$$

Let TP represent the number of true positives, TN represent the number of true negatives, FP represent the number of false positives, and FN represent the number of false negatives.

Parametric Analysis of the Proposed Work and Existing Techniques

Dataset Description

The Corel-2K dataset contains 1000 images that are classified as various groups, namely dinosaurs, buses, flowers, beaches, etc. Image size is 256 × 384 or 384 × 256. The total number of relevant ship, parachute, and signboard images in the database is 1000, 500, and 1500, respectively.

Table 1 shows the input and retrieved images using the proposed technique and its retrieval details. The number of relevant retrieved ship images are 630, 718, 840, and 980 while the relevant predicted ship images are 300, 250, 320, and 950 for CNN, KNN, LBP and RIV3-NET-EDBN methods, respectively. The number of relevant retrieved parachute images are200, 290, 390, and 480 while the relevant predicted parachute images are 100, 250, 330 and 460 for CNN, KNN, LBP, and RIV3-NET-EDBN methods respectively. The number of relevant retrieved signboard images is 666, 879, 1008 and 1400 while the relevant predicted signboard images are 400, 350, 420, and 1280 for CNN, KNN, LBP, and RIV3-NET-EDBN methods, respectively.

Table 2 shows the analysis of existing methods with the proposed methodology. The analysis of metrics reveals a complex performance landscape. Precision is generally low across all methods and images, ranging from 21.7% to 37.6%. In contrast, recall is consistently high, ranging from 65.2% to 90.9%, with RIV3-NET-EDBN performing best. The F1-scores are unusually low across all methods and images, which is unexpected given the Precision and Recall values and may indicate a calculation or reporting error. The accuracy shows a clear trend of improvement from CNN to RIV3-NET-EDBN, with RIV3-NET-EDBN achieving over 90% accuracy for all images.

Table 2: Analysis of Existing Methods with Proposed Methodology

Input image	*Methods*	*Precision (%)*	*Recall (%)*	*F1-Score (%)*	*Accuracy (%)*
Ship	CNN	21.7	65.2	0.326	69.12
	KNN	32.4	72.7	0.448	72.3
	LBP	34.6	77.2	0.478	80.6
	RIV3-NET-EDBN	34.9	84.3	0.494	84.8
Parachute	CNN	23.6	76.9	0.361	70.2
	KNN	27.8	81.5	0.415	74.1
	LBP	34.4	85.7	0.491	76.3
	RIV3-NET-EDBN	35.6	89.4	0.509	86.4
Signboard	CNN	24.8	89.1	0.388	72.9
	KNN	33.5	89.5	0.488	75.36
	LBP	34.8	86.3	0.496	79.4
	RIV3-NET-EDBN	37.6	90.9	0.532	89.8

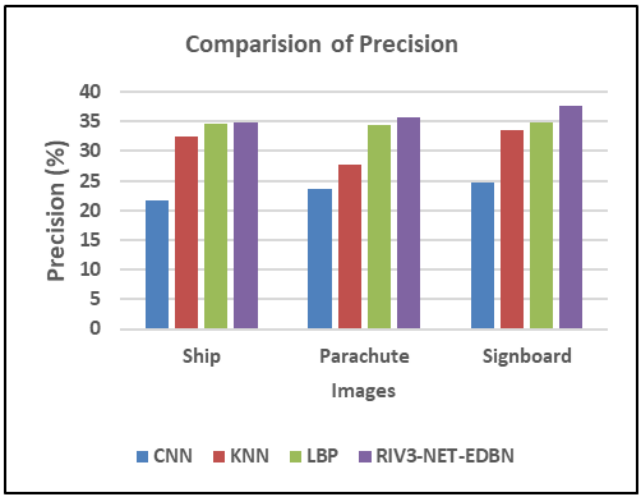


Figure 8: Comparison of Precision

Figure 8 shows that RIV3-NET-EDBN consistently outperforms other methods, demonstrating the highest precision in all categories. LBP follows closely behind, with high precision levels. KNN shows moderate effectiveness, particularly in the Signboard category, while CNN exhibits the lowest precision, suggesting less suitability for specific tasks.

Based on figure 9, the recall metrics across categories reveal RIV3-NET-EDBN's continued dominance, as it achieves the highest recall in all categories, underscoring its proficiency in identifying relevant instances. LBP and KNN both demonstrate strong recall performance, with LBP showing a slight edge in the Ship and Parachute categories.

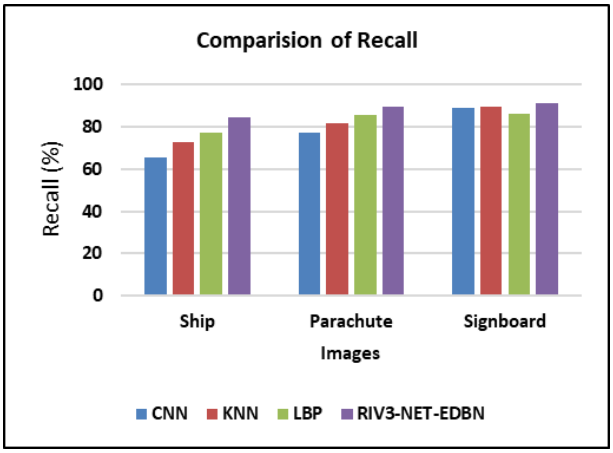


Figure 9: Comparison of Recall

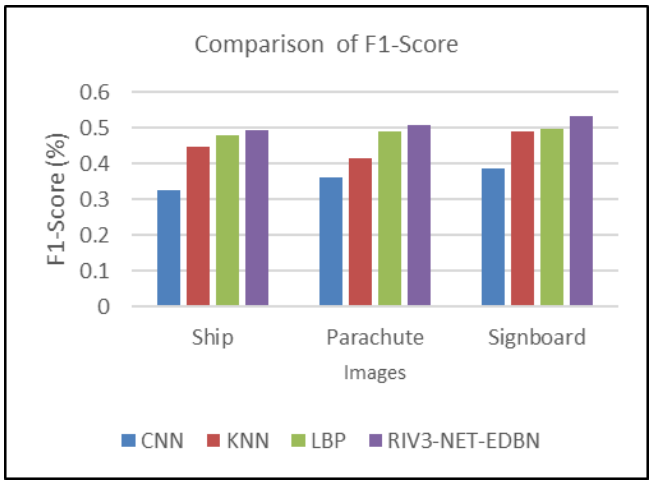


Figure 10: Comparison of F1- Score

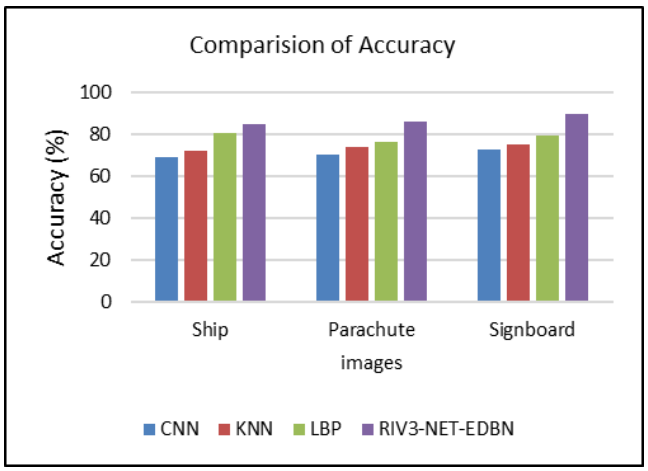


Figure 11: Comparison of Accuracy

CNN, while exhibiting the lowest recall for Ship and Parachute categories, manages to perform comparatively well in the Signboard category.

Figure 10 presents a comparison of the F1-Score between existing techniques and the proposed RIV-3 NET with EDBN, demonstrating an improvement in the F1-Score for the proposed methods.

The accuracy of the current method and the suggested one are compared in Figure 11. When it comes to picture retrieval, RIV3-NET with EDBN offers the best accuracy compared to other methods.

Conclusion

With a few notable exceptions, such as image segmentation, research into content-based image retrieval for commercial applications has had a minimal impact despite the existence of numerous sophisticated image retrieval methods. Selecting characteristics that represent people's actual interests is the unresolved problem. The RIV3-NET feature extraction and EDBN deep learning classification methods that are being suggested. For efficient feature extraction, RIV3-NET with EDBN and chi-square distance are utilised here. The experimental results are acquired for the image frame that contains the ship, parachute, and signboard. Previous work on KNN, CNN, and LBP is contrasted with these results. Retrieving the image frame images was 90% accurate using the suggested method.

References

Ali, F. (2020). Content-based image retrieval (CBIR) by statistical methods. *Baghdad Science Journal*, 17(2), 694–694.

Alrahhal, M., & Supreethi, K. P. (2019). Content-based image retrieval using local patterns and supervised machine learning techniques. *Amity International Conference on Artificial Intelligence (AICAI)*, 118–124.

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.

Alzu'bi, A. A., & Ramzan, N. (2017). Content-based image retrieval with compact deep convolutional features. *Neurocomputing*, 249, 95–105.

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.

Ashraf, R., Ahmed, M., Jabbar, S., Khalid, S., Ahmad, A., Din, S., & Jeon, G. (2018). Content-based image retrieval by using color descriptor and discrete wavelet transform. *Journal of Medical Systems*, 42(3), 1–12.

Datta, R., Joshi, D., Li, J., & Wang, J. Z. (2008). Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (CSUR)*, 40(2), 1–6.

Dharani, T., & Laurence Aroquiaraj, I. (2013). A survey on content-based image retrieval. International Conference on Pattern Recognition, *Informatics and Mobile Engineering*, 1(3), 485–490.

Hameed, I. M., Abdulhussain, S. H., & Mahmmod, B. M. (2021). Content-based image retrieval: A review of recent trends. *Cogent Engineering*, 8(1), *Article* 1927469.

Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 39(4), 3128–3137.

Kondylidis, N., Tzelepi, M., & Tefas, A. (2018). Exploiting TF-IDF in deep convolutional neural networks for content-based image retrieval. *Multimedia Tools and Applications*, 77(23), 30729–30748. https://doi.org/10.1007/s11042-018-5991-5

Koyuncu, H., Dixit, M., & Koyuncu, B. (2021). An analysis of content-based image retrieval. *International Advanced Research and Engineering Journal*, 5(1), 123–141.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

Li, Z., Tang, J., & Mei, T. (2018). Deep collaborative embedding for social image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9), 2070–2083.

ManickaChezian, R., & Janani, M. (2012). Content-based image retrieval system. *International Journal of Advanced Research in Computer Engineering & Technology*, 1(5), 554–604.

Qazanfari, H., AlyanNezhadi, M. M., & Nozari Khoshdaregi, Z. (2023). Advancements in content-based image retrieval: A

comprehensive survey of relevance feedback techniques. *arXiv preprint* arXiv:2312.10089.

Rehman, M., Iqbal, M., Sharif, M., & Raza, M. (2013). Content-based image retrieval: Survey. *World Applied Sciences Journal*, 19(3), 404–412.

Shi, X., Sapkota, M., Xing, F., Liu, F., Cui, L., & Yang, L. (2018). Pairwise based deep ranking hashing for histopathology image classification and retrieval. *Pattern Recognition*, 81, 14–22.

Sun, Y., Wang, X., & Tang, X. (2014). Deep learning face representation from predicting 10,000 classes. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 35(5), 1891–1898.

Wiggers, K. L., Britto, A. S., Heutte, L., Koerich, A. L., & Oliveira, L. E. S. (2018). Document image retrieval using deep features. *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 8(13), 1–8.

Yang, H. F., Lin, K., & Chen, C. S. (2018). Supervised learning of semantics-preserving hash via deep convolutional neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2), 437–451.

Yenigalla, S. C., Rao, K. S., & Ngangbam, P. S. (2023). Implementation of content-based image retrieval using artificial neural networks. *Engineering Proceedings*, 34(1), 25–33.

Zhang, C., Cheng, J., & Tian, Q. (2018). Multiview label sharing for visual representations and classifications. *IEEE Transactions on Multimedia*, 20(4), 903–913.

Zhang, C., Cheng, J., & Tian, Q. (2019). Multiview, few-labeled object categorization by predicting labels with view consistency. *IEEE Transactions on Cybernetics*, 49(11), 3834–3843.

Zhou, W., Li, H., Sun, J., & Tian, Q. (2018). Collaborative index embedding for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(5), 1154–1166.

Zhu, L., Shen, J., Xie, L., & Cheng, Z. (2017). Unsupervised visual hashing with semantic assistant for content-based image retrieval. *IEEE Transactions on Knowledge and Data Engineering*, 29(2), 472–486.