JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 185-201 (2021)
DOI: 10.6688/JISE.202101_37(1).0012

An Analysis of Two Novel and Efficient Deep Learning Models for Fast and Accurate Image Retrieval

SHIKHA BHARDWAJ[1], GITANJALI PANDOVE[2] AND PAWAN KUMAR DAHIYA[2]
[1]*Department of Electronics and Communication Engineering*
DCRUST, Murthal and UIET, KUK, Haryana, 136119 India
[2]*Department of Electronics and Communication Engineering*
DCRUST, Murthal, Haryana, 131039 India
E-mail: shikpank@yahoo.com; pandove_geet@yahoo.co.in; pawan.dahiya@gmail.com

Due to the presence of various limitations in traditional machine learning algorithms, the interest of the researcher community has been shifted towards Deep learning. In this paper, two Deep learning methods namely Deep belief network (DBN) and Stacked auto-encoder (SAE) have been analyzed for image retrieval task. But, in order to retrieve images from vast storehouses, more retrieval time is utilized. To solve this issue of retrieval time, two different indexing techniques namely Similarity-based indexing (SBI) and Cluster-based indexing (CBI) have been used. Thus, four models namely DBN-SBI, DBN-CBI, SAE-SBI and SAE-CBI have been developed and tested on two benchmark datasets, which are MIT-Vistex and INRIA-Holidays. Among these models, DBN-SBI obtains the highest results in terms of Precision, Recall and Retrieval time. Average precision of 98.45% and 86.53% with retrieval time of 0.035 seconds and 0.149 seconds has been obtained on Vistex and Holidays dataset respectively which is higher than many state-of-the-art related models.

Keywords: deep belief network, stacked auto-encoder, similarity-based indexing, cluster-based indexing, image retrieval

1. INTRODUCTION

The prevalence of digital images has led to an exceptional increase in the number of images present in image repositories available both online and offline. In order to retrieve images from such vast storehouses, an effective retrieval system is required. Retrieval of information can be done on the basis of type of data present in the repository. Data can be retrieved in the form of audio, video, text, images *etc.* Image retrieval is one of the eminent fields in the current era. This image retrieval can be done on the basis of manual annotation of text with images. But, this process suffers from various errors like misspellings, homonyms, synonyms *etc.* Therefore, retrieval of images based on varied image content like color, texture, shape, spatial information, edge *etc.*, is a brilliant option and the technique is popularly known by the name Content-based image retrieval (CBIR).

The main aim of CBIR system is to retrieve the most similar and relevant images based on the query image given by the user. But, the basic consideration in CBIR system is to obtain the maximum relevant information from the images in order to minimize the "Semantic Gap". The gap between low level image features and high-level human understanding is called semantic gap. Many types of techniques like Scale invariant feature transform (SIFT) [1], Speeded up robust features (SURF) [2], Local binary pattern (LBP)

Received December 8, 2019; revised August 6, 2020; accepted September 3, 2020.
Communicated by Flavia C. Delicato.

[3], Discrete wavelet transform (DWT) [4], Gray level co-occurrence matrix (GLCM) [5], Gabor [6, 7] *etc.* and many of their variants have been used earlier to recover the basic image information. Many machine learning algorithms like Support vector machine (SVM) [8], Extreme learning machine (ELM) [9], K-nearest neighbor (K-NN) [10], *etc.* can also be utilized with the basic feature extraction techniques so as to enhance the retrieval accuracy of the CBIR system. But, these techniques are not so effective in capturing the semantic information from the given images.

Therefore, the interest of the researcher community has been shifted from machine learning to its counterpart called Deep learning. The main aim of deep learning is to extract the results in a high level domain and thus capturing maximum semantic information from an image and has been utilized in numerous application areas like pattern recognition, voice recognition, computer vision *etc.* In deep learning, there are many techniques like Convolutional neural network (CNN), Restricted Boltzmann machine (RBM), Deep CNN and many structures with variants of each technique, which can be used in many prominent applications related to image processing.

But, yet another major issue of concern in CBIR system is the large time consumption during the retrieval of images. This happens so because each query image given by the user has to be matched with every image of the database in order to retrieve the similar images. Therefore, to minimize the retrieval time of images, many indexing techniques have been used. Indexing is a technique in which a database index of each image is maintained in a tabular format which can be used to access and find any image of that database [11]. Utilization of indexing techniques also leads to less storage requirements of the data and it can also be used as a sorting technique.

Therefore, based on the above facts and discussion, the main benefactions of this paper are as follows:

- To extract the features of the given image database by using two different and efficient deep learning models namely Deep belief network (DBN) and Stacked Auto-encoders (SAE) and an analytical comparison is done between the two techniques.
- In order to reduce the retrieval time of the developed image retrieval system, two different indexing techniques have been also tested and analyzed.
- Finally, the execution of the two utilized deep learning techniques based on two efficient indexing techniques have been done and the results have been evaluated by calculating Precision, Recall, Retrieval time, *etc.*
- All the analytical analysis has been done on two renowned and benchmark CBIR datasets namely MIT-Vistex [3] and INRIA-Holidays [12].

The remaining section of the paper is organized as follows: Related work has been given in Section 2, in Section 3, proposed methodology has been given. In Section 4, Results and discussions have been and lastly, Section 5 has been ended with the Conclusion.

2. RELATED WORK

This section dispenses varied deep learning models which have been used for the extraction of features from an image. A model based on deep CNN has been used by Tzelpi *et al.* [13] for feature extraction based on the deepest level of abstraction. The model has been retrained based on certain feedbacks. Another model based on a combination of CNN with Symmetric positive definite (SPD) matrix has been developed by Zhi gao *et al.* [14]. An end-to-end deep layer network has been designed through generation and transformation of SPD. But, in these systems, when the user enters the query image, the whole dataset has to be searched in order to retrieve the most relevant images and thus time complexity of the system increases, as no indexing technique has been utilized with the developed systems.

Deep learning techniques have been used in various applications related to interactive multimedia. One of the most eminent application is to sort or remove the duplicate images from vast storehouse of images. Deep CNN has been used in unsupervised mode [15] for near-duplicate detection of images [16]. Also, Mir-Flick near-duplicate dataset has been used as a ground truth. To detect near-duplicate images, CNN has been used by Yi Zhang *et al.* [17]. But, CNN suffers from certain pitfalls also. The information related to position and the orientation of an image is not encoded by CNN. All the information from the preceding neurons is routed to the same succeeding neuron irrespective of its positional information. In CNN, classification is also being done based on the prior predictions based on the given images.

In order to develop an effective image retrieval system, sometimes two or more than two techniques are combined to develop a hybrid system. Scale invariant feature transform (SIFT) has also been used as a popular feature extraction technique. This SIFT has been used in aggregation with CNN to develop a combined model and also both the techniques have been analyzed independently [18]. This model has been used to describe the given images in three levels which are point-level, scene-level and object-level. But, SIFT is based on histogram of gradients (HOG) and is computationally heavy with large mathematical computation. The same technique has been adopted by Zhu *et al.* [19] by aggregating deep low level convolutional attributes into a high dimensional feature vector for the required images.

Auto-encoders have also been used as a technique of deep learning. These auto-encoders have been used in various modes to retrieve the desired set of images. Different modes of auto-encoders like Siamese auto-encoder (SAE) [20], Separate classifier auto-encoder (SCA) [20] and Classifying auto-encoder (CAE) [20] have been used to retrieve fine-grained images. Auto-encoders have also been used in a comparative mode by comparing the performance of auto-encoders with other deep learning techniques like Drop out neural networks, Deep Boltzmann machine (DBM), Denoising auto-encoders [21], *etc.* Auto-encoder based on denoising technique has been used to reconstruct the original image from the noisy samples and is used in a combination with CNN to classify the data [22]. But, the developed systems with auto-encoders have huge retrieval time as there is no provision of maintaining any index or record of the extracted features. Also, these auto-encoders have also some limitations like, the retrieved images or results become blurry as the complexity of the images increases.

Both colored and gray scale images can be retrieved by using any CBIR system. For gray scale images, generally texture features are extracted. Texture describes different discernible patterns of an image. Radial mean local binary pattern (RMLBP) and Prototype data model (PDM) has also been used by sootodeh *et al.* [23] for texture feature extraction. Some variants of basic texture extraction techniques like Local neighborhood intensity pattern (LNIP) [24] has also been analyzed. Thus, depending upon the type of the image dataset, *i.e.*, colored or gray scale, different types of feature extraction techniques can be selected. Thus, the requirement of the present time is to extract the features of the image by using an effective deep learning technique. Also, the developed system should be based on an efficient indexing technique in order to fasten the retrieval time of the images. Thus, based on the above drawbacks of the developed systems, a technique is utilized in this paper which is an effective framework of a deep learning and indexing technique. Moreover, the proposed system has been analyzed on both colored and Gray scale image datasets.

3. PROPOSED METHODOLOGY

In this paper, an efficient image retrieval system is proposed which is based on a framework of combining a deep learning technique with an indexing technique. Deep belief network (DBN) and Stacked Auto-encoder (SAE) are the two deep learning techniques which have been used to extract the features of an image. These techniques are being compared in order to find out the best amongst the two. The two deep learning methodologies have also been checked on two varied indexing techniques namely Cluster-based indexing and Similarity-based indexing. An analytical comparison of these two indexing techniques have also been done in this paper. Formally, the methodology for the development of the proposed system can be divided into two main stages:

(1) Index Creation stage
(2) Index Prediction or Testing stage

In index creation stage, there are two main steps involved for the creation of indexes related to each image of the database, which are Feature extraction and Indexing. A brief description of feature extraction techniques namely DBN and SAE with both indexing methods is given below.

3.1 Deep Belief Network

Deep learning is an advanced and eminent technique based on the concepts of machine learning. The main aim of this technique to extract the results in a high level domain and thus capturing maximum semantic information from an image. A deep belief network (DBN) is a type of deep learning based neural network model which consists of multiple layers and these layers have connection amongst them but there exists no connection between the units of the layers [25]. The DBN consists of multiple and stacked layers of Restricted Boltzmann machine (RBM) and was first proposed by Hinton *et al.* [26]. Each RBM consists of hidden and visible layers. The learning in DBN takes place by configuring the 1st hidden layer and the visible layer into a single RBM. When the learning process is completed, the training of the hidden layers is done with RBM by using a new input as a value of the 1st hidden layer. Therefore, the output of the lower-layer RBM is used as an input for next-higher layer RBM, which also signifies that the hidden layer of previous RBM serves as the visible layers of the next RBM [27] as shown in Fig. 1.

3.2 Stacked Auto-Encoder

An auto-encoder is based on unsupervised learning and is a type of artificial neural network. The prime consideration of an auto-encoder is based on learning the representations of the data (encoding) by training the system in such a way, so as to ignore the present noise. With this reduction function, a side by side representation of the original input signal is also reconstructed [28]. Thus, it can be summed that it is a technique which is based on learning the input data and copies that input to its output. It consists of a hidden layer which depicts a code and that is used to represent an input. It has major two sections: an encoder section which maps the given input to the code and a decoder section which is used for the mapping the data from the code to reconstruct the original given input.

Thus, the basic auto-encoder is based on unsupervised learning and is a feed-forward neural network. It has one input layer, single output layer and one or more hidden layers which are used for the connection of the preceding and succeeding layers. The basic architecture of an auto-encoder is given in Fig. 2. There are various versions of auto-encoders like Stacked auto-encoders (SAE), Convolutional auto-encoder (CAE), De-noising auto-encoder (DAE), *etc.* which can be utilized for extracting the features of an image. Here, SAE is used for the extraction of image features which consists of multiple auto-encoder layers. The output of the previous layer is connected to the inputs of the next succeeding layer and due to this stacking, these auto-encoders avail all the benefits of deep learning techniques.

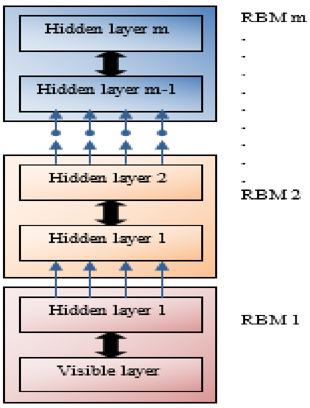

Fig. 1. A basic deep belief network model.

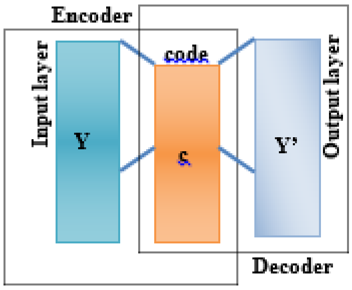

Fig. 2. A basic model of an auto-encoder.

3.3 Image Indexing (Cluster-based vs Similarity-based)

The two utilized techniques for indexing are cluster-based and Similarity-based. The description of both the techniques is as under:

(A) Cluster-based Indexing

Cluster based indexing approach is generally utilizes K-means clustering [29] algorithm. K-means is a partitioning algorithm which uses a two-step iterative process to find the user specified K clusters. If the set of observations ($y_1, y_2, y_3 \dots y_n$) are given then, the K-means clustering algorithm disintegrates the n observations into K parts ($K \leq n$) denoted by $S = (S_1, S_2, S_3 \dots S_k)$ in order to minimize the variance. Thus, the main objective of this algorithm [23] can be given as:

$$\min_{S_k} \sum_k \sum_{y \in S_k} \pi_y dist(y, m_k). \tag{1}$$

Here, in Eq. (1), S_k denotes k clusters, m_k denotes the centroid (centroid is the mean of the various cluster members) of S_k, $\pi_y > 0$ denotes the weight of y and *dist* is the distance function used. *K*-means is a two-phase iterative process and the steps can be given by means of an algorithm.

1 Initialize the initial value of *K* by using the criterion of 1% of the total dataset images.
2 Clusters are formed by connecting every input with the nearest mean.
3 Then, the arithmetic mean (Centroid) of each of the *K* clusters becomes the new mean and Steps 2 and 3 are repeated unless a convergence is reached and finally desired clusters are obtained.
4 End

Then, finally desired clusters are obtained and these clusters serve as indexes to different images of the dataset and index is formed for each cluster. This indexing technique is based on an initialization criterion of selecting 1% of the total dataset images as the initial value of *K*. According to this criterion cluster centres (*K*) of the order of 6 and 8 are obtained for Vistex and Holidays dataset respectively.

(B) Similarity-based Indexing

Similarity based indexing is a technique which is based on the usage of different similarity measures like Euclidean, Manhattan, Cosine, Minkowski, [30] *etc.* In this algorithm, the fusion of three attributes *i.e.*, color, texture and shape results in the formation of a fused feature matrix [31]. For similarity based indexing, the same criterion of 1% of the total dataset images is taken as the value of *K*. The complete procedure of this indexing can be explained by the means of an algorithm which is as follows:

1 Initially, the complete feature matrix of the obtained features is considered which is of size $M \times N$, where M represents total images in the dataset and N is the number of extracted features.
2 Row-wise sorting or grouping of the obtained feature matrix is done by arranging the obtained feature vector in ascending order and a new re-ordered matrix is obtained.
3 Disintegration of the newly obtained feature matrix into *K* parts (1% of the total images in the dataset) as per the chosen criterion.
4 Calculation of mean or average for each of the *K* obtained part and it becomes the respective cluster centre.
5 Each image is assigned to its nearest cluster centre by using a specific similarity metric technique.
6 Finally, required cluster centres are obtained and each image during the query session is matched with only the respective cluster centre and not with all images of the dataset.
7 Now, indexes for each image is formed using the above technique.
8 End

Also, these deep learning models, viz. SAE and DBN in their basic form does not possess the characteristics of clustering and indexing. But, by using these two algorithms in amalgamation with SAE and DBN, four different models are formed which can perform feature extraction, clustering and indexing in contrast to SAE and DBN which can only be utilized for feature extraction.

Now, the main steps for Index creation stage can be summarized as follows:

Feature Extraction Stage: It is the first step of the index creation stage. Here, the two deep learning based neural network models namely DBN and SAE are utilized independently to extract the high level features of an image and two separate feature vectors are thus obtained.

Clustering and Index Creation Stage: The two independent feature vectors obtained by using DBN and SAE are applied as an input to two indexing algorithms namely Cluster-based indexing and Similarity-based indexing. After using these two indexing techniques, suitable *K* clusters have been formed which serve as indexes for each of them individually and thus index for each cluster is obtained. The architecture for this part of the proposed system is given in Figs. 3 (a) and (b) respectively for DBN and SAE.

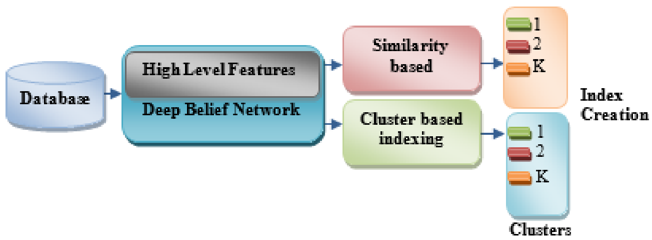

Fig. 3. (a) Proposed model for DBN based on two indexing techniques.

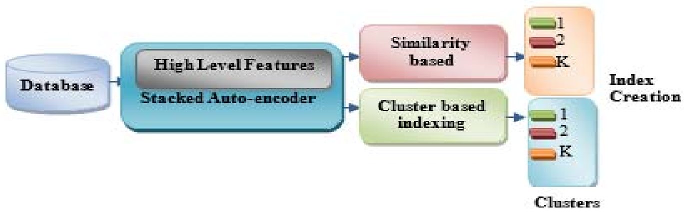

Fig. 3. (b) Proposed model for SAE based on two indexing techniques.

(2) Index prediction/Testing stage: In this stage, the indexes which are created by using two different indexing approaches in the previous section are utilized to retrieve the desired images by using a specific distance metric in spite of matching each and every image of the database. The main subsections of this stage are as follows:

Query formulation stage: It is the first stage of this section. The user enters the query image to test the working of the proposed system. Here, each and every image is utilized as a query image.

Feature extraction stage: Again, the features of the query image are extracted using the two techniques namely DBN and SAE and two independent feature vectors are obtained for each of them.

Index Prediction and Similarity matching stage: In this stage, the matching of the query image is performed in two levels. Firstly, the matching of two obtained feature vectors are done with the already created clusters and the best cluster is chosen for each of them. In the second level, by using a specific distance metric, matching of the obtained feature vector is performed with images in that selected cluster and finally top *N* images are retrieved based on the minimal distance. The same procedure is adapted for both of the indexing techniques. The basic architecture depicting this part is given in Fig. 4.

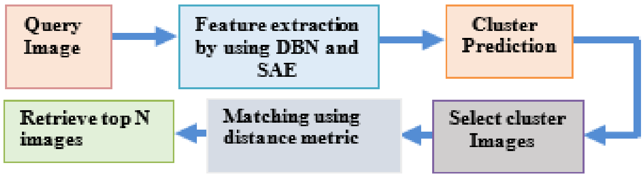

Fig. 4. Proposed model for DBN and SAE using query formulation stage.

4. EXPERIMENTAL ANALYSIS AND RESULTS

To analyze the retrieval efficiency, two benchmark datasets namely MIT-Vistex and INRIA-Holidays have been utilized. All the experiments are performed in MATLAB R2017a, core i3 processor, 4 GB memory, 64-bit windows.

A brief description of these datasets is given below and sample images from each dataset is shown in Fig. 5.


Fig. 5. Sample images from Vistex (Gray scale) and Holidays (Colored) datasets.

1st Dataset: MIT-Vistex: It is a dataset of gray scale images and it consists of 40 initial images and finally, a database of 640 images [32] is created by partitioning the original 512×512 image into 128×128 sixteen non-overlapping images (40×16). Now finally, the obtained dataset contains 8 categories and each category has 80 images.

2nd Dataset: INRIA Holidays: It is a dataset which contains images related to personal holiday scenes [33] like nature, fire effects, *etc.* This dataset contains 1491 images which are further divided into 500 categories, where each image corresponds as a query image and remaining 991 images are utilized as training images. Here, the number of images in each category is not fixed.

4.1 Evaluation Metrics

There are different evaluation parameters which can be utilized for calculating the capability of a particular CBIR system. But, the prominent metrics are Precision and Recall [34, 35].

$$Precision(P_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \quad (2)$$

$$Recall(R_i) = \frac{Number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ retrieved} \quad (3)$$

4.2 Retrieval Performance Analysis

In order to verify the retrieval performance of the proposed system, various experiments have been conducted. The average precision obtained by using both the indexing techniques for DBN and SAE on Holidays dataset is given in Tables 1 and 2. Similarly, the average precision obtained over Vistex dataset by varying the number of retrieved images from 10 to 50 with the utilization of both, similarity-based indexing and cluster-based indexing for DBN and SAE both are given in Tables 3 and 4.

Table 1. Average precision (%) on Holidays dataset based on similarity-based indexing.

Database	Technique	Average Precision (%) by using Similarity-based Indexing
Holidays	DBN-SBI	86.53
	SAE-SBI	79.05

Table 2. Average precision (%) on Holidays dataset based on cluster-based indexing.

Database	Technique	Average Precision (%) by using Cluster-based Indexing
Holidays	DBN-CBI	74.71
	SAE-CBI	72.36

Table 3. Average precision (%) on Vistex dataset based on similarity-based indexing.

Database	No. of Images	Average Precision (%) using Similarity-based Indexing	
MIT-Vistex		Deep Belief Network	Stacked auto-encoder
	10	98.45	97.2
	20	96.25	95.3
	30	90.83	92.5
	40	85.31	81.25
	50	77.25	70.75

Table 4. Average precision (%) on Vistex dataset based on cluster-based indexing.

Database	No. of Images	Average Precision (%) using Cluster-based Indexing	
MIT-Vistex		Deep Belief Network	Stacked auto-encoder
	10	96.25	94.3
	20	94.32	92.1
	30	89.75	87.63
	40	82.81	81.25
	50	73.25	70.75

In Tables 1-4, the results depict that among the two utilized deep learning models, DBN has superior results as compared to SAE. Also, from the two utilized indexing techniques, similarity-based indexing has higher results in comparison to cluster-based indexing. This is so because, DBN has more number of hidden layers and can retrieve more similar images as compared to SAE and by using similarity-based indexing, the whole utilized dataset can be divided into varied parts depending upon the choice of the user. Thus, Deep belief network based on similarity based indexing (DBN-SBI) has higher results than any other combination, *i.e.* (DBN-CBI, SAE-SBI and SAE-CBI).

In addition, Wilcoxon signed rank test has also been done to compare the four developed models. The tests essentially calculate the difference between sets of pairs and analyzes these differences to establish if they are statistically and significantly different from one another. Firstly, DBN-SBI is compared with SAE-CBI and then DBN-CBI is compared with SAE-SBI to find the difference between given pairs. The results obtained are given in Table 5.

In order to enhance the comparison, box plot charts have been shown in Figs. 6 and 7. The box plot chart in Fig. 6 depicts the deviation in precision and recall values by using the four techniques for Vistex dataset and Fig. 7 depicts the box plot chart for deviation in precision from the average value by using Holidays dataset. In Fig. 6, each technique has been shown two times on x-axis by T1, T2, T3 and T4, one each for precision and recall values.

Table 5. Results of Wilcoxon signed rank test using four techniques on Vistex dataset.

Pairs used	Value of Z	Value of p	Value of W
DBN-SBI and SAE-CBI	$Z = -1.7838$	$p = .03754$ significant at $p < .05$	$W = 10$ Critical value of W at $N = 10$ is 10
DBN-CBI and SAE-SBI	$Z = -2.4973$	$p = .03754$ significant at $p <. 05$	$W = 3$ Critical value of W at $N = 10$ is 10

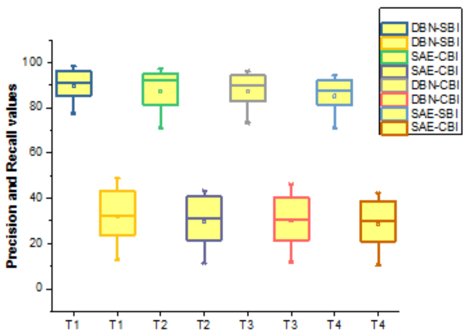

Fig. 6. Box-plot on Vistex dataset.

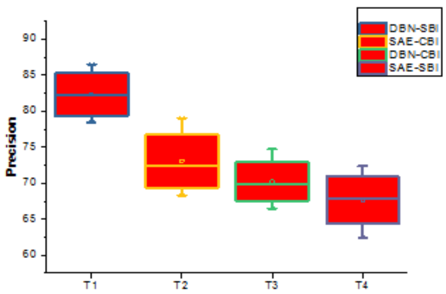

Fig. 7. Box-plot on Holidays dataset.

There are different distance metrics which have been utilized for calculating the similarity between a query image and complete database images. Both the indexing techniques have been evaluated with varied distance metrics. The results obtained by using varied distance metrics [36, 37] on similarity-based indexing on both the datasets are given in Tables 6 and 7. Some of the prominent distance metrics are:

$$Distance_{Euclidean} = \sqrt{\sum_{j=1}^{n} (|I_j - D_j)^2|} \tag{4}$$

$$Distance_{Minkowski} = \left[\sum_{j=1}^{n} (|I_j - D_j)^{1/P}|\right] \tag{5}$$

The results in Tables 6 and 7 shows that among all the utilized distance metrics, Euclidean has the highest results because it is based on weighted and normalized features of an image. The number of images for each cluster for Vistex dataset is given in Table 8.

Table 6. Average precision (%) on various distance metrics on Vistex dataset.

Dataset	Technique	Average Precision (%) on various distance metrics				
		Euclidean	Manhattan	Cosine	Jaccard	Minkowski
Vistex	DBN	98.45	81.25	80.5	83.75	77.25
	SAE	97.2	73	70.75	43	70.75

Table 7. Average precision (%) on various distance metrics on Holidays dataset.

Dataset	Technique	Average Precision (%) on various distance metrics				
		Euclidean	Manhattan	Cosine	Jaccard	Minkowski
Holidays	DBN	86.53	82.5	82.82	59.12	79.5
	SAE	79.05	76.25	74.5	50.71	73.16

Table 8. No. of images present in each cluster based on Vistex dataset.

Cluster ($K = 8$)	Record count (No. of images/cluster)
1	49
2	174
3	81
4	246
5	56
6	34

In Table 8, the results depict that for Vistex dataset, the whole dataset has been divided into 6 clusters ($K = 6$) and the record count represents the number of images present in each cluster. Then, if a query image is given by the user, its best matching cluster is obtained based on the minimum distance and finally, only the images present in that particular cluster are matched with the given query image to retrieve the top N images. The important parameters of the model are given in Table 9.

Table 9. Various working parameters of the proposed system.

Deep belief network	Activation function = Restricted Boltzmann Machine (RBM) (2) RBM 1(Max. Epoch 10), RBM 2 (Max. Epoch 10)
Stacked Auto-encoder	Activation function = Spasity regularizer No. of layers = 2 Epoch = 5
Similarity-based Indexing	$K = 6$ (For Vistex), $K = 8$ (For Holidays)

4.3 Time Performance Analysis

Since the proposed hybrid descriptor is based on indexing, therefore the retrieval time [38] is indeed an important aspect to be considered. This section describes the retrieval time of the two utilized techniques namely DBN and SAE based on similarity-based indexing on both Vistex and Holidays dataset and the comparison is given in Table 10.

Table 10. Retrieval time comparison between two utilized techniques.

Technique	Dataset	Retrieval time (in seconds)
DBN	Vistex	0.073
SAE	Vistex	0.105
DBN	Holidays	0.149
SAE	Holidays	0.168

The results in Table 10 shows that DBN has less retrieval time as compared to SAE on both the utilized datasets. In order to validate the obtained results, the developed system has been compared with many state-of-the-art techniques in terms of retrieval time and average precision. In terms of retrieval time, the developed system uses less time to retrieve the desired images. This can be confirmed by focusing on Figs. 8 (a) and (b) for Vistex and Holidays dataset respectively.

The results in Figs. 8 (a) and (b) shows that DBN-SBI (PM) has less retrieval time as compared to the related systems based on both the datasets.

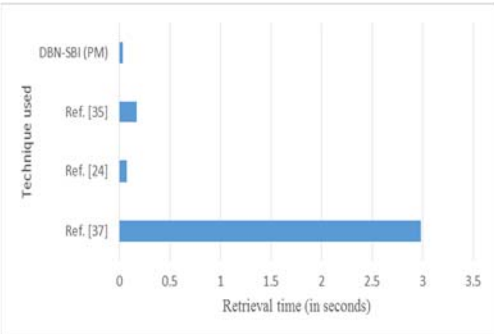

Fig. 8 (a). Retrieval time comparison on Vistex dataset.

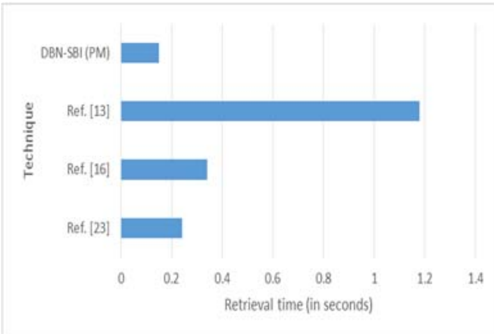

Fig. 8 (b). Retrieval time comparison on Holidays dataset.

4.4 State-of-the-Art Comparison to Other Techniques

From all the experiments conducted over the two utilized deep learning models namely DBN and SAE on both the indexing techniques, it can be concluded that Deep belief network-Similarity based indexing (DBN-SBI) combination has the highest results. Therefore, this DBN-SBI model has been compared to many-state-of-the-art techniques based on both MIT-Vistex and INRIA-Holidays dataset and the comparison based on both these datasets is given in Tables 11 and 12 for Vistex and Holidays dataset respectively.

Table 11. State-of-the-art comparison on Vistex dataset.

Dataset	Technique	Average precision (%)
Vistex	Ref. [24]	90
	Ref. [3]	86.53
	Ref. [35]	89.9
	Ref. [37]	97.5
	Ref. [38]	93.5
	Ref. [7]	92.14
	DBN-SBI(PM)	98.45

Table 12. State-of-the-art comparison on Holidays dataset.

Dataset	Technique	Average precision (%)
Holidays	Ref. [24]	65.71
	Ref. [16]	69.4
	Ref. [34]	71.6
	Ref. [13]	74
	Ref. [15]	77
	Ref. [19]	85.7
	DBN-SBI(PM)	86.53

The results in Tables 11 and 12 depict that the proposed model (PM) based on the exceptional combination of DBN-SBI has superior results as compared to many of the latest compared techniques on both Vistex and Holidays. Most of the compared techniques based on both the datasets does not employ any indexing technique for the fast retrieval of images. Moreover, some of the compared systems does not even employ any deep learning model to retrieve the desired images. But, the proposed system is a magnificent combination of deep learning with indexing and can retrieve the higher level of semantic information from the utilized datasets.

Individual Graphical user interface (GUI) [12, 39] has been designed for each dataset based on the top N retrieval. To show the retrieval accuracy, the top N images retrieved by using independent GUI for each dataset has been given in Figs. 9 (a) and (b).

From Figs. 9 (a) and (b), the results show that based on the given query image by the user, the similar and relevant top N images have been retrieved by using the proposed DBN-SBI system. For Holidays dataset, top 5 images have been retrieved as there are maximum 6 images per category. And for Vistex dataset top 10 images have been retrieved.

Fig. 9. (a) Retrieval results of top 5 images from Holidays dataset.

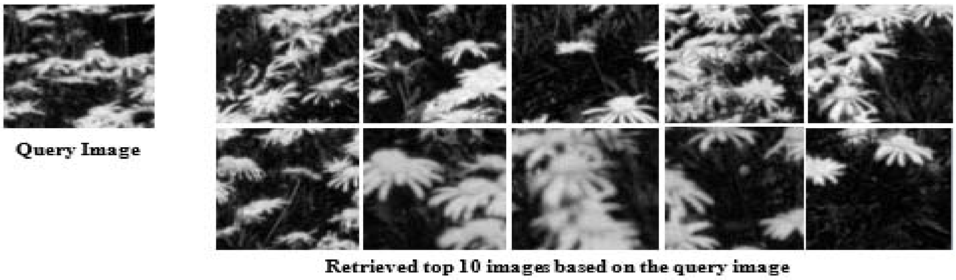

Fig. 9. (b) Retrieval results of top 10 images from Vistex dataset.

5. CONCLUSIONS

In this paper, two Deep learning models namely DBN and SAE have been investigated for their usage in image retrieval systems. These two models have been tested on two diverse indexing techniques namely Similarity-based indexing (SBI) and Cluster-based indexing (CBI). Thus, based on these techniques, four different models have been created and tested namely DBN-SBI, DBN-CBI, SAE-SBI and SAE-CBI. And from these models, it can be concluded that DBN-SBI model has the highest results. To validate the given statements, various simulation results have been provided in terms of Precision, Recall, Retrieval time and the provided results have been compared to many existing and latest techniques. The future work will be concentrated on developing an efficient CBIR system with advanced versions of DBN and SAE with more number of hidden layers and to develop the system for practical applications like finger print detection, Industrial application *etc.* with more superior indexing techniques.

REFERENCES

1. G. Feng, Y. Jie, C. Yilei, and B. Yao, "Research on image detection and matching based on SIFT features," in *Proceedings of the 3rd International Conference on Control and Robotics Engineering*, 2018, pp. 130-134.
2. A. Narpat, B. Manoj, and M. C. Bala, "Face recognition system based on SURF and LDA technique," *International Journal of Intelligent Systems and Applications*, Vol. 2, 2016, pp. 13-19.
3. V. Naghashi, "Co-occurrence of adjacent sparse local ternary patterns : A feature descriptor for texture and face image retrieval," *International Journal of Light Electron Optics*, Vol. 157, 2018, pp. 877-889.
4. N. S. T. Sai, P. Ravindra, S. Shailesh, and N. Bhushan, "Truncated DCT and decomposed DWT SVD features for image retrieval," *Procedia Computer Science*, Vol. 79, 2016, pp. 579-588.
5. S. Bhardwaj, G. Pandove, and P. K. Dahiya, "An intelligent multi-resolution and co-occuring local pattern generator for image retrieval," *EAI Endorsed Transactions on Scalable Information System*, Vol. 6, 2019, pp. 1-12.
6. S. Pathan and P. C. Gopalakrishna, "A pixel processing approach for retinal vessel extraction using modified Gabor functions," *Program Artificial Intelligence*, 2017,

DOI 10.1007/s13748-017-0134-4, pp. 1-14.

7. E. Walia and V. Verma, "Boosting local texture descriptors with Log-Gabor filters response for improved image retrieval," *International Journal of Multimedia Information Retrieval*, Vol. 5, 2016, pp. 173-184.
8. H. Khalifi, W. Cherif, A. E. Qadi, and Q. Youssef, "Query expansion based on clustering and personalized information retrieval," *Program Artificial Intelligence*, Vol. 8, 2019, pp. 241-251.
9. C. Jan, W. Jian, and M. Jacek, "Review and performance comparison of SVM- and ELM-based classifier," *Neurocomputing*, Vol. 128, 2014, pp. 507-516.
10. J. Rejito and D. Setiana, "Image indexing using color histogram and *K*-means clustering for optimization CBIR," *International Journal of Computer Applications Technology and Research*, Vol. 6, 2017, pp. 405-409.
11. S. P. Meshram, A. D. Thakare, and P. Santwana, "Hybrid swarm intelligence method for post clustering content based image retrieval," *Procedia Computer Science*, Vol. 79, 2016, pp. 509-515.
12. K. S. Govindan and V. K. Arun, "A hybrid deep learning architecture for latent topic-based image retrieval," *Data Science and Engineering*, Vol. 3, 2018, pp. 166-195.
13. M. Tzelepi and A. Tefas, "Image communication deep convolutional image retrieval: A general framework," *Signal Processing and Image Communication*, Vol. 63, 2018, pp. 30-43.
14. Z. Gao, Y. Wu, X. Bu, T. Yu, J. Yuan, and Y. Jia, "Learning a robust representation via a deep network on symmetric positive definite manifolds," *Pattern Recognition*, Vol. 92, 2019, pp. 1-12.
15. L. P. Valem, D. Carlos, and G. Pedronette, "Unsupervised selective rank fusion for image retrieval tasks," *Neurocomputing*, Vol. 377, 2020, pp. 182-199.
16. L. Morra and F. Lamberti, "Benchmarking unsupervised near-duplicate image detection," *Expert Systems with Applications*, 2019, pp. 313-326.
17. Y. Zhang, Y. Zhang, J. Sun, H. Li, and Y. Zhu, "Learning near duplicate image pairs using convolutional neural networks," *International Journal of Performability Engineering*, Vol. 14, 2018, pp. 168-177.
18. K. Yan, Y. Wang, D. Liang, T. Huang, and Y. Tian, "CNN vs. SIFT for image retrieval: Alternative or complementary," in *Proceedings of the 24th ACM International Conference on Multimedia*, 2016, pp. 407-411.
19. J. Zhu, J. Wang, S. Pang, W. Guan, Z. Li, Y. Li, and X. Qian, "Co-weighting semantic convolutional features for object retrieval," *Journal of Visual Communication and Image Represenation*, Vol. 62, 2019, pp. 368-380.
20. T. Portenier, Q. Hu, P. Favaro, and M. Zwicker, "Fine-grained retrieval with autoencoders," in *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Vol. 5, 2018, pp. 85-95.
21. Q. Xu, S. Jiang, and W. Huang, "Feature fusion based image retrieval using deep learning," *Journal of Information and Computational Science*, Vol. 12, 2015, pp. 2361-2373.
22. S. S. Roy, M. Ahmed, and M. A. Akhand, "Noisy image classification using hybrid deep learning methods," *Journal of Information and Communication Technolog*y, Vol. 17, pp. 233-269.

23. M. Sotoodeh, M. R. Moosavi, and R. Boostani, "A novel adaptive LBP-based descriptor for color image retrieval Mahmood," *Expert Systems with Applications*, Vol. 127, 2019, pp. 342-352.
24. P. Banerjee, A. K.Bhunia, A. Bhattacharyya, P. P. Roy, and S. Murala, "Local neighborhood intensity pattern – A new texture feature descriptor for image retrieval," *Expert Systems with Applications*, Vol. 113, 2018, pp. 100-115.
25. J. M. Kim, *et al.*, "A deep belief network and dempster-shafer-based multiclassifier for the pathology stage of prostate cancer," *Journal of Healthcare Engineering*, Vol. 2018, 2018, pp. 1-8.
26. M. Krizhevsky and G. E. Hinton, "Using very deep autoencoders for content-based image retrieval," in *Proceedings of the 19th European Symposium on Artificial Neural Networks*, 2011, pp. 1-11
27. J. Yu and X. Yan, "Active features extracted by deep belief network for process monitoring," *ISA Transactions*, Vol. 84, 2019, pp. 247-261.
28. P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of JMLR Workshop and Conference Proceedings*, 2012, pp. 37-50.
29. C. T. Chang, L. Jim, and J. Mu, "A fuzzy *K*-means clustering algorithm using cluster center displacement," *Journal of Information Science and Engineering*, Vol. 27, 2011, pp. 995-1009.
30. J. Cao, Z. Wu, J. Wu, and W. Liu, "Towards information-theoretic *K*-means clustering for image indexing," *Signal Processing*, Vol. 93, 2013, pp. 2026-2037.
31. Y. Mistry, D. T. Ingole, and M. D. Ingole, "Content based image retrieval using hybrid features and various distance metric," *Journal of Electrical System and Information Technol*ogy, Vol. 2016, 2017, pp. 1-15.
32. S. Fadaei, P. Amirfattahi, and M. R. Ahmadzadeh, "New content-based image retrieval system based on optimised integration of DCD, wavelet and curvelet features," *IET Image Processing*, Vol. 11, 2017, pp. 89-98.
33. S. Bhardwaj, G. Pandove, and P. K. Dahiya, "An extreme learning machine-relevance feedback framework for enhancing the accuracy of a hybrid image retrieval system," *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 6, 2020, pp. 15-27.
34. L. Liao, Y. Zhao, S. Wei, and Y. Zhao, "Improving image similarity estimation via global distance distribution," *Neurocomputing*, Vol. 333, 2019, pp. 381-394.
35. M. Verma and B. Raman, "Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval," *Journal of Visual Communication and Image Representation*, Vol. 32, 2015, pp. 224-236.
36. L. K. Pavithra and T. S. Sharmila, "An efficient framework for image retrieval using color, texture andand edge features," *Computer and Electrical Engineering*, 2017, pp. 1-14.
37. R. Khemchandani and P. Saigal, "Color image classification and retrieval through ternary decision structure based multi-category," *Neurocomputing*, Vol. 165, 2015, pp. 444-455.
38. Y. R. Charles and R. Ramraj, "A novel local mesh color texture pattern for image retrieval system," *International Journal of Electronics and Communication*, Vol. 70, 2016, pp. 225-233.

39. E. Mehdi, E. Aroussi, and N. E. Houssif, "Content-based image retrieval approach using color and texture applied to two databases (Coil-100 and Wang)," in *Proceedings of the 9th International Conference on Soft Computing and Pattern Recognition*, 2017, pp. 49-59.

Shikha Bhardwaj received her B.Tech degree and M.Tech degree in ECE. Her major research areas are image processing, artificial intelligence, machine learning *etc.*

Gitanjali Pandove received her Ph.D. degree in ECE. She is working as a Professor in ECE Department, in DCRUST, Murthal, Haryana, India. Her research areas are optical communication, digital image processing, power systems, *etc.* She has more than 70 publications in various international reputed journals

Pawan Kumar Dahiya received his Ph.D. degree in ECE. He is working as a Professor in ECE Department, in DCRUST, Murthal, Haryana, India. His research areas are digital and embedded system design, ANPR, VANETs, soft computing techniques, image processing, Internet of Things (IoT), *etc.* He has more than 75 publications in various national and international reputed journals.