Content-Based Image Retrieval Using AutoEmbedder

Md. Mohsin Kabir[1], Adit Ishraq[1], Kamruddin Nur[2], and M. F. Mridha[1]
[1]Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka, Bangladesh
[2]Department of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
Email: {mdmkabi, iamaditishraq}@gmail.com, kamruddin@aiub.edu, firoz@bubt.edu.bd

***Abstract*—Content-Based Image Retrieval (CBIR) technique attempts to retrieve relevant query images from the extensive repositories of images. With the advancements of the internet and multimedia technology, images have increased at a significant rate. Retrieving similar pictures from a vast database has always been an arduous task where CBIR techniques are helpful. However, similar images retrieval efficiency improvement is a common problem with the available CBIR techniques due to inadequate feature sets. This paper proposes a novel CBIR technique using a Deep Convolutional Neural Network (DCNN)-based AutoEmbedder. With this novel approach, this study attempt to map the higher dimensional features into relevant clusterable embeddings with k-means clustering to cluster the relevant images. The architecture is evaluated using the Corel10K and CIFAR-10 datasets, and the average precision and recall value is used to evaluate the architecture's performance. The proposed model's significance is that it outperforms the existing CBIR techniques presented in experimental results.**

***Index Terms*—Content-Based Image Retrieval (CBIR), Deep Convolutional Neural Network (DCNN), AutoEmbedder, K-means clustering**

I. Introduction

One of the exoteric media formats to understand and interact with the human community is images. Image datasets have regarded histrionic deployment since the advancements of digital cameras and digital platforms. Processing image data has become an open field of interest successfully in image research due to the swift improvement of the internet and multimedia terminals. In addition, the complexity and diversity of images have increased the difficulty for users and automated systems to retrieve expected pictures from vast image libraries. Therefore, an accurate and efficient image retrieval system is required to extract such obligate knowledge from extensive multimedia storage.

Usually, people use three ways to retrieve images: text, content, and semantic-based [1], [2]. Image retrieval using text is the traditional text-based approach that describes each image using a text notation, and the relevant image is retrieved by searching with the keyword [3]. These keywords are annotated on images to carry out the text-based search for the expected photos. Nevertheless, this approach requires manual labeling and annotation, which becomes complicated and ambiguous for a database of gigantic size. Consequently, text-based approach is replaced by (CBIR) [4]. A CBIR algorithm intends to measure the association between images by incorporating low-level characteristics or features. These features characterize each image's content and retrieve images related to the search query image [5]. The CBIR approach extracts characteristics from the query image and then searches for these extracted features in the entire database of the stored images. A feature vector represents every image in the database via CBIR. The feature vector for the query image is also computed after receiving the query mage. Finally, the CBIR technique compares it with the stored feature vectors then return a similar vector to the query image. CBIR has been an important research topic in digital media since the 1990s [6]. Fig. 1 depicts a typical CBIR system.

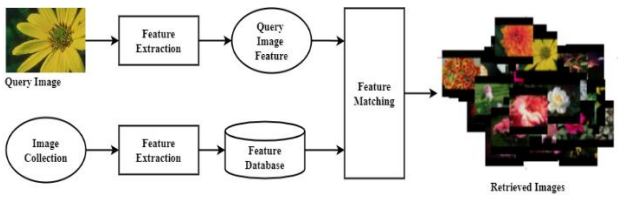

Figure 1. Typical content-based image retrieval system.

Recently, several Semantic-Based Image Retrieval (SBIR) architectures have been suggested [1]. Traditional CBIR approaches [7], [8] performed well dealing with uncomplicated image data. The performance starts to degrade whenever images vary from basic scenery to more semantically complicated images, such as images of animals or different objects (with natural background) [9].

The reason for accuracy reduction solely depends on the feature extraction techniques employed in traditional CBIR methods. The canny edge method, grey level co-occurrence matrix, discrete wavelet transform, and canny edge histogram are less accurate in extracting higher-dimensional features from a complex set of images [10]. The similarity computing algorithm, such as the metaheuristic algorithm used in the CBIR technique not being advanced, can be another reason. A dynamic and intelligent image retrieval approach needs to be introduced to boost the accuracy of a CBIR technique on an extensive

Manuscript received December 13, 2021; revised April 24, 2022.

doi: 10.12720/jait.13.3.240-248

database. With the advancements of machine learning, Deep Neural Networks (DNNs) are solving many real-life problems. The Deep Convolutional Neural Network (DCNN) makes it easier to extract features from images. Shiv Ram Dubey [11] presented how deep learning-based techniques improve image retrieval on content-based in the last decades of 2011 to 2020. An extensive survey featured different state-of-art methods, including networks, supervision, descriptor (binary, real-value) type, and retrieval (cross-modal, semantic, etc.) type. The author showed the best performance of Autoencoder, generative adversarial, reinforcement learning networks and discovered accuracy enhancement of image retrieval by using the objective function, feature learning, and quantization techniques. In this paper, a Deep CNN (DCNN)-based AutoEmbedder is used to sieve the features from photos. A semi-supervised model called AutoEmbedder is a Deep Neural Network (DNN) embedding model that downsamples high dimensional data to clusterable embedding points [12]. Furthermore, the features are sieved for a query image, and relevant cluster assignment is performed utilizing the k-means clustering algorithm [13].

The paper presents a novel CBIR technique based on the recently developed AutoEmbedder [12], which downsamples the higher-dimensional images to clusterable embedding points for clustering. We also present the improved version of the Deep-CBIR technique based on the AutoEmbedder.

The major contributions of this paper are presented as follows:

- We propose a novel DCNN-based approach to implement the content-based image retrieval system. Specifically, we implement the AutoEmbedder strategy to generate clusterable embedding of images.
- We implement the k-means clustering algorithm to find relevant matches of the feature vectors. A comparison between different CBIR approaches is described.
- We validate the proposed CBIR system with multiple data sets; compare the results with other traditional CBIR systems and found that the proposed CBIR system outperformed other existing state-of-the-art CBIR techniques.

The remainder of the study is structured as follows: The past and present research study regarding CBIR is presented in Section II. Here we have presented the relevant research of this domain. The proposed deep CBIR technique is exhibited and illustrated in Section III. This section mathematically and theoretically explained the proposed technique. However, the next Section IV presents the evaluation of the proposed technique on the Corel-10K and CIFAR-10 datasets. Lastly, Section V concludes the article.

II. Related Work

Current advancements in Deep Learning (DL) have obtained magnificent innovations in several domains, particularly in computer vision, where artificial intelligence has advanced beyond human capability. Lately, a transformation has been perceived in characteristic design from hand to learning-based engineering following the appearance of modern DL [14], [15]. In the last few years, researchers have proposed numerous deep learning-based supervised, unsupervised, semi-supervised CBIR systems.

Amid the beginning efforts, Krizhevsky *et al.* [16] have applied a deep Autoencoder to outline the pictures to short binary codes for CBIR. Kang *et al.* [17] introduced a deep multi-view hashing to produce the code for CBIR. Wu *et al.* [18] acknowledged the reoccurring pre-trained stacked denoising Autoencoders. Wang *et al.* [19] suggested a deep ranking architecture that receives the correlation metric instantly from images. Further, Zhang *et al.* [20] trained a DCNN model to optimize CBIR's discriminative image features and hash functions. Then, a supervised Deep Hashing Network (DHN) is introduced by Zhu *et al.* [21] to determine essential feature representation by establishing the quantization error, and Cao *et al.* [22] offered a Deep Quantization Network (DQN), which is significantly related to the DHN architecture.

After DQN, Cao *et al.* [23] introduced HashNet deep architecture to produce the hash code by a preservation scheme. Su *et al.* [24] applied the greedy hash by transferring the gradient for the hash coding layer in a CBIR system. Fu *et al.* [25] proposed CNN and Support Vector Machine (SVM)-based CBIR systems. The authors applied CNN to generate in-depth features from the images and employed SVM to train a hyperplane to find similar images. Ng *et al.* [26] extracted convolutional features from diverse convolutional neural network layers and received Vector of Locally Aggregated Descriptors (VLAD) encoding for representing features with vectors. The authors revealed that higher layers with upper scales give better performance on image retrieval tasks. Wan *et al.* [27] proposed a deep learning architecture for CBIR. The authors found that the neural network can imitate the brain of humans. The main advantage of deep learning is that it can extract hierarchical features for many semantic abstractions and efficiently improve CBIR's accuracy. However, most of these CBIR deep learning architectures were based on hash coding. Apart from the hash coding technique, several techniques have been recommended based on the texture and color combination extraction such as local and global color and texture feature extraction.

Das *et al.* [28] proposed an extraction method using image binarization to improve the retrieval of images. The authors tested the system with two public datasets containing 3688 images. The number of features was 12 dimensional, and recall and precision were utilized to evaluate the technique. Hiremath *et al.* [29] introduced an image retrieval method by combining local color, texture, and global shape features. The color and shape are used as the local descriptors, and shape information is used as global descriptors. The model represents a way for determining salient points based on color saliency that provides a feature for image retrieval. Ashraf *et al.* [30]

introduced an image representation and feature extraction method using a Bandelet Transform (BT) to retrieve images from a massive dataset. The authors utilized artificial neural networks for the retrieval of photos. The authors used three public datasets (Coil, Corel, and Caltech 101) to evaluate the model and used precision and recall to assess efficiency.

Irtaza *et al.* [31] proposed a Genetic Classifier Comity Learning (GCCL) by combining ANN with SVMs. The authors showed that using multiple classifier can potentially improve the performance of CBIR systems. This method enhanced performance and outperformed several state-of-the-art techniques. Qi *et al.* [32] introduced a cross-category label propagation algorithm to improve the image retrieval approach. The model can generate inter-category learning at the instance level among the source and target classes. Pun *et al.* [33] proposed an architecture for CBIR using rectangular segmentation. The authors created a significant function to adjust for the various image locations and enhance image retrieval performance. Zeng *et al.* [34] introduced a Local Structure Descriptor (LSD) to retrieve color images. Shape, texture, and color are combined by LSD to retrieve images. The Corel dataset was used to evaluate the model, and the result convinced the superiority of the architecture. Zhou *et al.* [35] researched on color image retrieval where the authors proposed new architecture by combining two forms of histograms, Local Directional Pattern (LDP) and color signature. The authors evaluated the model on two public datasets (Wang's and Corel) and compared the result with other cutting-edge techniques.

Nazir *et al.* [36] proposed a CBIR method for color and texture feature combinations. The authors used Color Histogram (CH) to extract the color information and utilized the Discrete Wavelet Transform (DWT) and Edge Histogram Descriptor (EDH) for the extraction of texture features. The authors used the Corel dataset to evaluate the model and used precision and recall to assess efficiency. Jhanwaret *et al.* [37] proposed a CBIR architecture utilizing the Motif Co-occurrence Matrix (MCM). MCM combines the information for both color and texture. By using fuzzy logic, Ahmed *et al.* [38] further enhanced the image retrieval performance. The image is described with a Fuzzy Attributed Relational Graph (FARG) which indicates every object in the picture. The Human Vision System (HSV) is used to compute texture and color attributes. Huang and Dai [39] developed a fuzzy matching algorithm that incorporated energy distribution pattern strings that quickly retrieved the desired images. This method relied on the textural similarity and used features from a mix of energy distribution pattern strings and sub-band gradient vectors. Mezzoudj *et al.* [40] proposed a Spark framework for ContentBased Image Retrieval (S-CBIR) that used Tachyon to store feature vectors at fast memory speed. The author sped up the process using two criteria (1) image Indexation by MapReduce distributed model and (2) image retrieval by parallel k Nearest Neighbor (k-NN) search with and without cache. Experimental results on the ImageNet dataset highlighted the highest runtime compared to the previous methods. Yashwant and Vijayshri [41] proposed a histopathology image segmentation algorithm on CBIR to enhance imagery disease diagnostic and grading system. Three steps followed the approach: pre-processing using histogram and multi radial kernel; identifying and extracting nuclei region; executing nuclei segmentation. Performance analysis evaluated on five HI and one multi-organ histopathology images datasets and results showed a 6% upgrade in F1 score, 23% improvement in Hausdorff distance, and 7% in their proposed combined parameter named WCF.

Feng *et al.* [42] introduced a Global Correlation (GC) vector, Directional Global Correlation (DGC) vector, and Global Correlation (GC) descriptor for image retrieval. The authors validated their findings using the Corel-5K and Corel-10K datasets, and they measured efficiency using recall and precision. Ali *et al.* [43] demonstrated a CBIR approach using the scale-invariant and robust features. These features improve the efficiency of the model. The authors used the Corel 1500 and 1000 datasets to evaluate the model. Liu *et al.* [44] proposed a Multi-texton Histogram (MTH) to retrieve images that incorporate the histogram and co-occurrence matrix. The MTH descriptor represents the spatial correlation of color and texture orientation and finds the query image from a massive dataset. Rao *et al.* [45] introduced a CBIR system with two steps. First, the authors used hierarchical clustering to filter the pictures, and then they submitted the clustered images to the RBFN network to efficiently obtain the query image. Cheun-Horng Lin *et al.* [46] proposed three-step image retrieval systems called CTCHIRS. In the first two steps, color and texture features are described as CCM, and the third is called Color Histogram for K-Means (CHKM). Finally, a sequential forward selection approach is introduced for the selection of features.

These existing techniques mostly focused on low-dimensional features for content based image retrieval. With the use of histogram or Gabor filter techniques can not extract high-dimensional feature resulting in poor results. Hence, the DCNN is introduced for the image retrieval concept.

III. Methodology

In this paper, a feasible CBIR using AutoEmbedder is introduced where the higher-dimensional features of the images have been considered. K-means clustering algorithm is used to identify the similarity between the query image and dataset images.

A. Proposed Algorithm

The proposed CBIR technique is divided into two segments, training AutoEmbedders and retrieving similar images. In the first segment, higher-dimensional image features of the dataset images are extracted by the DCNN. The second segment carries out the image retrieval task by performing a k-means clustering algorithm. The overview of the proposed architecture is described in Fig. 2.

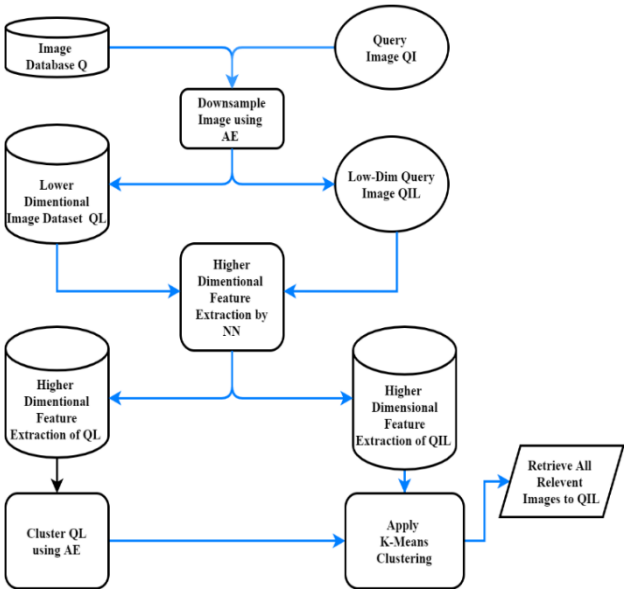

Figure 2. The figure illustrates the proposed CBIR architecture. First, the images are downsampled using AE. Then neural networks are applied for extracting high dimensional features. Finally, AE clustered images, and k-means retrieved the relevant images.

B. *Proposed Algorithm*

The overall architecture of the proposed CBIR technique is divided into two modules. The first module extracts the higher dimensional features of image datasets using the deep neural network, and AutoEmbedder [12] is used to downsample the higher dimensional data to clusterable embedding points. This module is further divided into three layers. Downsampling image to lower dimension using AEm, extracting higher-dimensional features using neural network, and using AEm cluster the lower dimensional images by higher dimensional extracted features. The second module is divided into five layers. The first three layers work the same as the previous module, and the fourth layer finds the similarity between the query image and image datasets. Finally, the last layer applies the k-means clustering algorithm to reach the point that is most relevant to the query image. The two modules can be expressed as Algorithm 1.

In the Algorithm 1, we have used a modified version of AutoEmbedder (AEm). AutoEmbedder is a novel embedding technique where the higher-dimensional data is downsampled to clusterable embedding points [12]. AEm relates to traditional classifier deep neural network architecture with a pairwise loss reduction method. In feature learning, to compute pairwise constraint loss, AEm employs a Siamese network Architecture. The training process is semi-supervised. The embedding function is generated using the DNN architecture, which is then trained using the SNN architecture. The learned embedding function is then used to convert higher dimensions data into low dimensional embedding points on which clustering may be conducted.

Algorithm 1: Proposed Deep-CBIR Algorithm

Input: Subset of the dataset for training X, AutoEmbedder model M, Number of iterations *epochs*, Training batch per iteration *batchSize*, Distance hyper parameter α, Query images Q

Result: Retrieve images of the query image

initialize an AutoEmbedder models m; Build a siamese network S with AutoEmbedder m as identical subnetwork;

$iter \leftarrow 0$;

while $iter < epochs$ **do**

  initialize two empty input data set I and I^0;

  initialize an empty target output set Y;

  $b \leftarrow 0$;

  while $b < batchSize$ **do**

    Randomly make a boolean choice on 0.5 probability;

    if *choice is true* **then**

      select two random data input x_i and x_j containing must link constraint $c_{ij} == 0$;

      append x_i to I, x_j to I^0 and c_{ij} to Y;

    else

      select two random data input x_i and x_j containing must link constraint $c_{ij} == \alpha$;

      append x_i to I, x_j to I^0 and c_{ij} to Y;

    $b = b + 1$;

  train siamese network S with inputs I, I^0 and Y;

  $iter = iter + 1$;

Generate embedding points of X, using trained AutoEmbeder AE;

Apply k-mean to generate embedding point of X;

for *each* $q \in Q$ do

  Generate embedding point of the query image q;

  Apply k-means to find relevant image of the coresponding q;

AEm relates to traditional classifier deep neural network architecture with a pairwise loss reduction method. In feature learning, to compute pairwise constraint loss, AEm employs a Siamese network Architecture. The training process is semi-supervised. The embedding function is generated using the DNN architecture, which is then trained using the SNN architecture. The learned embedding function is then used to convert higher dimensions data into low dimensional embedding points on which clustering may be conducted.

IV. Evaluation

In this section, at first, we explain two datasets that are used for evaluation. Then, we describe the evaluation metrics we have used along with the experimental setup of the AutoEmbedder based image retrieval system. Finally, the experimental results are explained.

A. *Dataset*

Several datasets of images are used for the CBIR problem, such as Corel, CIFAR-10 [47], NUS-WIDE [48], MNIST [49], SVHN [50], etc. Among them, the Corel and CIFAR-10 datasets are most renowned for evaluating CBIR systems. To elongate the efficiency of the proposed

architecture, the Corel dataset has 10,800 images used in terms of image retrieval. The Corel dataset has a couple of versions, Corel-10K, Corel 5K, Corel 1K, etc. Corel 1K has 1000 images grouped in 10 semantic classes like Food, beach, flowers, elephants, buildings, mountains, buses, africa, dinosaurs, flowers, and horses [51]. Corel-10K has 10800 images of several groups such as Buildings, Food, Mountains, Beach, Elephants, Buses, Horses, Flowers, Africa, Dinosaurs, etc. [52]. CIFAR-10 has 6000 images per class in 10 classes. It has semantic categories like Frog, automobile, truck, horse, bird, cat, deer, airplane, dog, horse, ship, and automobile. There are 60000 pictures in the CIFAR-10 dataset, 10000 test images, and 50000 training images.

It has semantic categories like Frog, automobile, truck, horse, etc. The CIFAR-10 has 6000 images per class in 10 classes. bird, cat, deer, airplane, dog, horse, ship, and automobile. There are 60000 pictures in the CIFAR-10 dataset, 10000 test images and 50000 training images.

B. Evaluation Metric

The performance of the intended CBIR approach is evaluated by determining the abundance of related images recovered in response to query images. Evaluation metrics such as precision and recall rates are also used where precision measures the capacity of the proposed system to recover pictures identical to the query image. The recall rate is considered a true positive rate to evaluate the power of the proposed system comparing all similar photos with the quantity of recovered relevant photos in the database. The equations of recall and precision rates are defined as follows:

$$Precision = \frac{Number\ of\ similar\ images\ retrived}{Total\ Number\ of\ retrieved\ images} \quad (1)$$

$$Recall = \frac{No.\ of\ similar\ images\ retrived}{Total\ No.\ of\ similar\ images\ in\ the\ database} \quad (2)$$

C. Experimental Setup

We have used Google Colab [53] for the purpose of experiments. We have used TensorFlow [54] DCNN [55] and AutoEmbedder [12] to achieve the system implementation using Keras [56]. To evaluate of our model, we used the Corel10k dataset's [57] images with the dimension of 120×120×3. In total, 10,0000 images were used for the evaluation purpose. The standard training-validation split was 85%-15%.

D. Experiments and Comparisons on Corel-10K

This article presented an AutoEmbedder-based image retrieval system that can return similar images from a massive database. The proposed system reduced the dimension of the images from the distinct dataset to lower and cluster them to their relevant classes. For the evaluation of the proposed model, first, we applied the algorithm to Corel-10K dataset. Fig. 3 displays the clustered points after experimenting with the system on the Corel-10K dataset. Here, we employed the euclidean distance of the k-means clustering algorithm to find all the relevant images to the query image. The Fig. 3 clearly presents 10 different cluster of Corel-10K dataset.

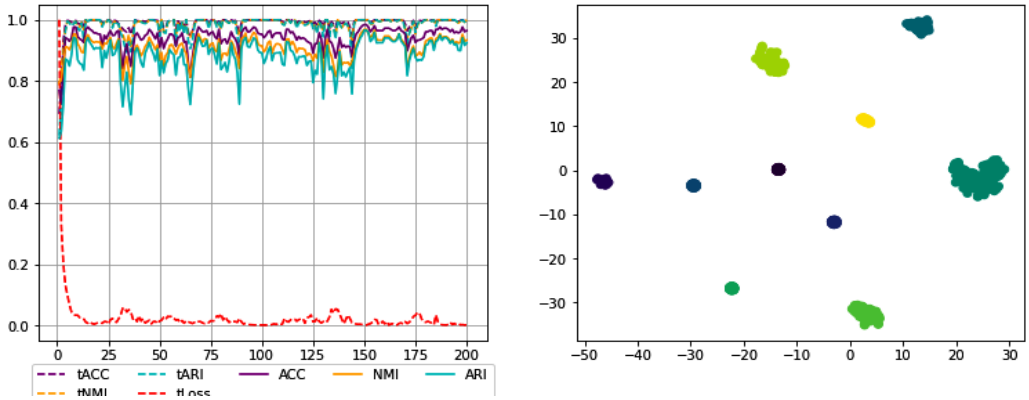

Figure 3. The figure illustrates the clustered points of the Corel-10K dataset.

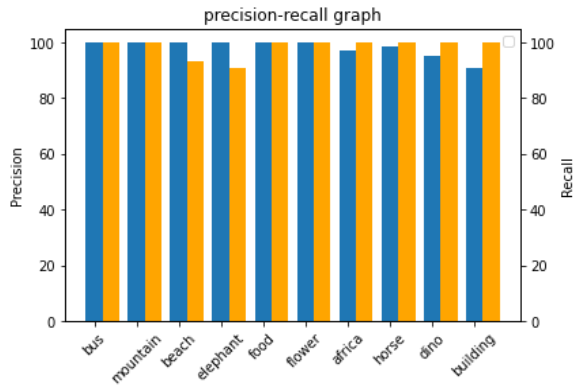

Figure 4. The figure presents the precision and recall scores of proposed model on the Corel-10K dataset.

Typically, the effectiveness of a machine learning model can be measured by the standard evaluation metrics - precision and recall. Similarly, the precision and recall metrics indicate the effectiveness and ability of the proposed system to retrieve images identical to the query image. Fig. 4 displays the precision and recall respectively for the Corel-10K dataset. The Fig. 4 illustrates that, for bus, mountain, food, flower classes, the proposed model gives the precision and recall score of 1.00. For other classes, the model also demonstrated higher precision and recall as shown in Fig. 4.

However, while modeling CBIR algorithms for higher-dimensional data, the performance often degrades due to higher complexity. To overcome this, we applied AutoEmbedder, a semi-supervised system that reduces the

higher dimensional feature by keeping the necessary features [12]. These essential features help to retrieve relevant images after applying k-means clustering algorithm. The backbone of the proposed algorithm, Siamese Neural Network (SNN), gets a pair of inputs and forms a pair of outputs. First, a pairwise loss is measured by computing the pairwise distance of the SNN-AutoEmbedder included embedding. This loss is additionally decreased using the conventional backpropagation method simultaneously with an optimization function. Thus, the models perform higher accuracy with modified AutoEmbedder and k-means clustering-based architecture.

We also compare the outcomes of our model with existing state-of-the-art CBIR systems proposed by Alsmadi *et al.* [58], Madhavi *et al.* [59], Ashraf *et al.* [30], Youssef [60], Rao *et al.* [61], and Lin *et al.* [46]. Table I presents the comparison between the proposed system and the existing systems in terms of average precision values on the Corel-10K dataset. Similarly, Table II shows the comparison between the proposed approach and the existing systems in terms of average recall values on the Corel-10K dataset. The experimental results demonstrated that the proposed approach has the maximum precision and recall rates in all categories compared to similar systems. In addition, the proposed system gives the highest average precision and recall values of 0.981 and 0.98, respectively, which is greater than all other systems, as shown in Table I and Table II.

TABLE I. THE TABLE PRESENTS PRECISION VALUES OF THE PROPOSED MODEL AND DIFFERENT POPULAR ARCHITECTURES ON COREL-10K DATASET

Class	Alsmadi *et al.* (2017) [46]	Madhavi *et al.* (2016) [47]	Ashraf *et al.* (2015) [16]	Youssef (2012) [48]	Rao *et al.* (2011) [49]	Lin *et al.* (2009) [32]	Proposed method
Bus	0.96	0.84	0.95	0.92	0.89	0.88	**1.00**
Mountain	0.82	0.81	0.75	0.74	0.51	0.52	**1.00**
Beach	0.90	0.89	0.70	0.64	0.53	0.54	**1.00**
Elephant	0.83	0.72	0.80	0.78	0.57	0.65	**1.00**
Food	0.87	0.87	0.75	0.81	0.69	0.73	**1.00**
Flower	0.96	0.91	0.95	0.95	0.89	0.89	**1.00**
Africa	0.83	0.82	0.65	0.64	0.56	0.68	**0.97**
Horse	0.96	0.95	0.90	0.95	0.78	0.80	**0.98**
Dino	0.99	0.82	1.00	0.99	0.98	0.99	**0.95**
Building	0.75	0.63	0.75	0.70	0.61	0.54	**0.90**
Average	0.888	0.83	0.82	0.81	0.70	0.72	**0.981**

TABLE II. THE TABLE SHOWS THE RECALL VALUES OF THE PROPOSED MODEL AND DIFFERENT POPULAR ARCHITECTURES ON COREL-10K DATASET

Class	Alsmadi *et al.* (2017) [46]	Madhavi *et al.* (2016) [47]	Ashraf *et al.* (2015) [16]	Youssef (2012) [48]	ElAlami *et al.* (2011) [50]	Lin *et al.* (2009) [32]	Proposed method
Bus	0.75	0.73	0.19	0.18	0.11	0.12	**1.00**
Mountain	0.75	0.73	0.15	0.15	0.22	0.21	**1.00**
Beach	0.81	0.80	0.14	0.13	0.19	0.19	**0.93**
Elephant	0.58	0.53	0.16	0.16	0.15	0.14	**0.91**
Food	0.62	0.60	0.15	0.16	0.13	0.13	**1.00**
Flower	0.66	0.64	0.19	0.19	0.11	0.11	**1.00**
Africa	0.73	0.70	0.13	0.13	0.15	0.14	**1.00**
Horse	0.85	0.84	0.18	0.19	0.13	0.13	**1.00**
Dino	0.75	0.72	0.20	0.20	0.09	0.10	**1.00**
Building	0.62	0.58	0.15	0.14	0.18	0.17	**1.00**
Average	0.71	0.69	0.16	0.16	0.14	0.14	**0.98**

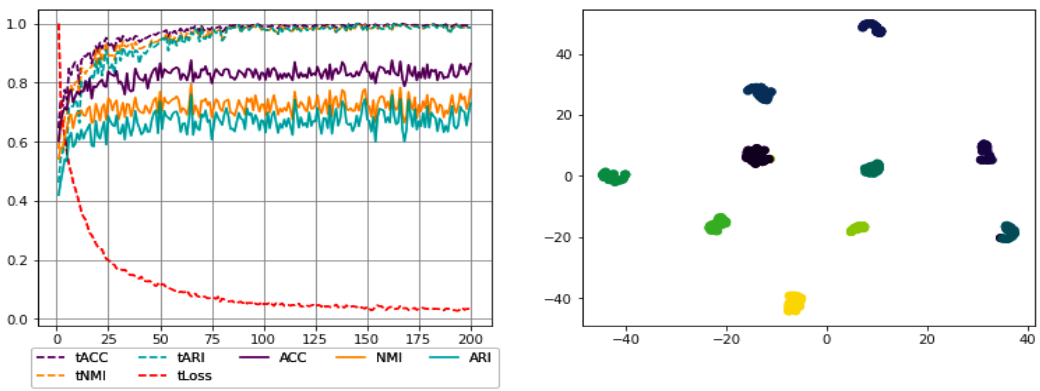

Figure 5. The figure illustrates the clustered points of CIFAR-10 dataset.

E. Validation with CIFAR-10 Dataset

In addition, to justify the outcomes of the proposed model on the Corel-10K dataset, we also evaluate the proposed model with CIFAR-10 dataset. At first, we cluster the query images, which is shown in Fig. 5. We can see from the image that all 10000 images of the CIFAR-10 validation set or query image set are separated into ten distinct clusters. The input shape of CIFAR-10 is 32×32×3. The training split has 50000 images, and the validation split has 10000 images to find the precision and recall score.

Finally, the precision and recall scores of each category of the CIFAR-10 dataset are shown in Fig. 6. The figure shows that the precision and recall scores of every category are much higher, which led us to conclude that the proposed model presents state-of-the-art results on the image retrieval task.

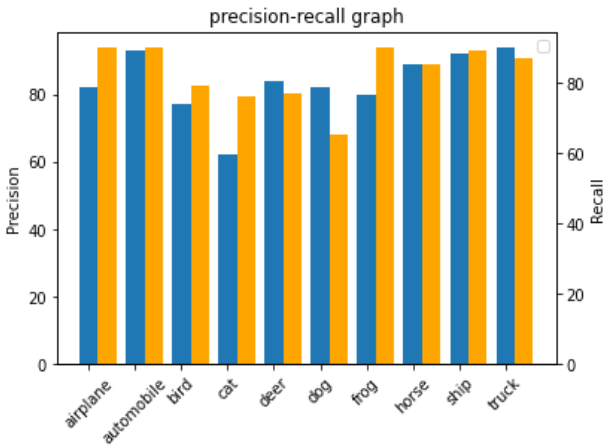

Figure 6. The figure presents the precision and recall scores of proposed model on the CIFAR-10 dataset.

V. Conclusion

This paper proposes an effective CBIR technique using AutoEmbedder and Deep CNN to retrieve similar images from vast image datasets. The proposed CBIR model extracts higher-dimensional features from the image using the neural network. Then, all the images are clustered using the AutoEmbedder algorithm. Finally, by following the query image, relevant images are retrieved by k-means clustering algorithm by measuring euclidean distance. The proposed system is trained and evaluated by two benchmark datasets Corel-10K and CIFAR10. The experiment results revealed that the proposed method outperforms all the existing CBIR techniques for Corel-10K and CIFAR-10 dataset.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

The research is conducted under the supervision of M. F. Mridha and K. Nur. The coding part is programmed by A. Ishraq and the initial draft is written by M. Kabir with A. Ishraq. The final version is written by K. Nur and M. F. Mridha.

References

[1] Y. Liu, D. Zhang, G. Lu, and W. Y. Ma, "A survey of content-based image retrieval with high-level semantics," *Pattern Recognition*, vol. 40, no. 1, pp. 262-282, 2007.

[2] Y. Liu, D. Zhang, and G. Lu, "Region-based image retrieval with high-level semantics using decision tree learning," *Pattern Recognition*, vol. 41, no. 8, pp. 2554-2570, 2008.

[3] I. K. Sethi, I. L. Coman, and D. Stan, "Mining association rules between low-level image features and high-level concepts," in *Data Mining and Knowledge Discovery: Theory, Tools, and Technology III*, International Society for Optics and Photonics, 2001, vol. 4384, pp. 279-290.

[4] S. Jaime-Castillo, J. M. Medina, and D. Sanchez, "A system to perform cbir on x-ray images using soft computing techniques," in *Proc. IEEE International Conference on Fuzzy Systems*, 2009, pp. 1314-1319.

[5] G. Carneiro, A. B. Chan, P. J. Moreno, and N. Vasconcelos, "Supervised learning of semantic classes for image annotation and retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 394-410, 2007.

[6] A. A. Radwan, B. A. A. Latef, A. M. A. Ali, and O. A. Sadek, "Using genetic algorithm to improve information retrieval systems," *World Academy of Science and Engineering Technology*, vol. 17, no. 2, pp. 6-13, 2006.

[7] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. C. Jain, and C. F. Shu, "Virage image search engine: an open framework for image management," in *Storage and Retrieval for Still Image and Video Databases IV*, International Society for Optics and Photonics, 1996, vol. 2670, pp. 76-87.

[8] C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equitz, "Efficient and effective querying by image content," *Journal of Intelligent Information Systems*, vol. 3, no. 3-4, pp. 231-262, 1994.

[9] R. Srihari, Z. Zhang, and A. Rao, "Image background search: Combining object detection techniques with content-based image retrieval (cbir) systems," in *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1999, pp. 97-101.

[10] M. K. Alsmadi, "Content-based image retrieval using color, shape and texture descriptors and features," *Arabian Journal for Science and Engineering*, vol. 45, pp. 1-14, 2020.

[11] S. R. Dubey, "A decade survey of content based image retrieval using deep learning," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.

[12] A. Q. Ohi, M. Mridha, F. B. Safir, M. A. Hamid, and M. M. Monowar, "Autoembedder: A semi-supervised dnn embedding system for clustering," *Knowledge-Based Systems*, p. 106190, 2020.

[13] K. Alsabti, S. Ranka, and V. Singh, "An efficient k-means clustering algorithm," working paper, 1997.

[14] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: An astounding baseline for recognition," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 806-813.

[15] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, 2020.

[16] A. Krizhevsky and G. E. Hinton, "Using very deep autoencoders for content-based image retrieval," *ESANN*, vol. 1, p. 2, 2011.

[17] Y. Kang, S. Kim, and S. Choi, "Deep learning to hash with multiple representations," in *Proc. IEEE 12th International Conference on Data Mining*, 2012, pp. 930-935.

[18] P. Wu, S. C. Hoi, H. Xia, P. Zhao, D. Wang, and C. Miao, "Online multimodal deep similarity learning with application to image retrieval," in *Proc. the 21st ACM International Conference on Multimedia*, 2013, pp. 153-162.

[19] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, B. Chen, and Y. Wu, "Learning fine-grained image similarity with deep ranking," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1386-1393.

[20] R. Zhang, L. Lin, R. Zhang, W. Zuo, and L. Zhang, "Bit-scalable deep hashing with regularized similarity learning for image retrieval

and person re-identification," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4766-4779, 2015.
[21] H. Zhu, M. Long, J. Wang, and Y. Cao, "Deep hashing network for efficient similarity retrieval," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, no. 1, 2016.
[22] C. Yue, M. Long, J. Wang, Z. Han, and Q. Wen, "Deep quantization network for efficient image retrieval," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 3457-3463.
[23] Z. Cao, M. Long, J. Wang, and P. S. Yu, "Hashnet: Deep learning to hash by continuation," in *Proc. the IEEE International Conference on Computer Vision*, 2017, pp. 5608-5617.
[24] S. Su, C. Zhang, K. Han, and Y. Tian, "Greedy hash: Towards fast optimization for accurate hash coding in cnn," in *Proc. the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 806-815.
[25] R. Fu, B. Li, Y. Gao, and P. Wang, "Content-based image retrieval based on cnn and svm," in *Proc. 2nd IEEE International conference on computer and communications*, 2016, pp. 638-642.
[26] J. Y. Ng, F. Yang, and L. S. Davis, "Exploiting local features from deep networks for image retrieval," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 53-61.
[27] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in *Proc. the 22nd ACM International Conference on Multimedia*, 2014, pp. 157-166.
[28] R. Das, S. Thepade, S. Bhattacharya, and S. Ghosh, "Retrieval architecture with classified query for content based image recognition," *Applied Computational Intelligence and Soft Computing*, vol. 2016, 2016.
[29] P. Hiremath and J. Pujari, "Content based image retrieval using color, texture and shape features," in *Proc. 15th International Conference on Advanced Computing and Communications*, 2007, pp. 780-784.
[30] R. Ashraf, K. Bashir, A. Irtaza, and M. T. Mahmood, "Content based image retrieval using embedded neural networks with bandletized regions," *Entropy*, vol. 17, no. 6, pp. 3552-3580, 2015.
[31] A. Irtaza, S. M. Adnan, K. T. Ahmed, A. Jaffar, A. Khan, A. Javed, and M. T. Mahmood, "An ensemble based evolutionary approach to the class imbalance problem with applications in cbir," *Applied Sciences*, vol. 8, no. 4, p. 495, 2018.
[32] G. J. Qi, C. Aggarwal, Y. Rui, Q. Tian, S. Chang, and T. Huang, "Towards cross-category knowledge propagation for learning visual concepts," in *Proc. CVPR*, 2011, pp. 897-904.
[33] C. M. Pun and C. F. Wong, "Content-based image retrieval using rectangular segmentation," *International Journal of Computers*, vol. 2, no. 1, pp. 72-79, 2008.
[34] Z. Zeng, "A novel local structure descriptor for color image retrieval," *Information*, vol. 7, no. 1, p. 9, 2016.
[35] J. X. Zhou, X. D. Liu, T. W. Xu, J. H. Gan, and W. Q. Liu, "A new fusion approach for content based image retrieval with color histogram and local directional pattern," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 4, pp. 677-689, 2018.
[36] A. Nazir, R. Ashraf, T. Hamdani, and N. Ali, "Content based image retrieval system by using hsv color histogram, discrete wavelet transform and edge histogram descriptor," in *Proc. International Conference on Computing, Mathematics and Engineering Technologies*, 2018, pp. 1-6.
[37] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, "Content based image retrieval using motif cooccurrence matrix," *Image and Vision Computing*, vol. 22, no. 14, pp. 1211-1220, 2004.
[38] H. Aboulmagd, N. El-Gayar, and H. Onsi, "A new approach in content-based image retrieval using fuzzy," *Telecommunication Systems*, vol. 40, no. 1-2, p. 55, 2009.
[39] P. W. Huang and S. Dai, "Image retrieval by texture similarity," *Pattern Recognition*, vol. 36, no. 3, pp. 665-679, 2003.
[40] S. Mezzoudj, A. Behloul, R. Seghir, and Y. Saadna, "A parallel content-based image retrieval system using spark and tachyon frameworks," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 2, pp. 141-149, 2021.
[41] Y. Kurmi and V. Chaurasia, "Content-based image retrieval algorithm for nuclei segmentation in histopathology images," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3017-3037, 2021.
[42] L. Feng, J. Wu, S. Liu, and H. Zhang, "Global correlation descriptor: A novel image representation for image retrieval," *Journal of Visual Communication and Image Representation*, vol. 33, pp. 104-114, 2015.
[43] N. Ali, K. B. Bajwa, R. Sablatnig, S. A. Chatzichristofis, Z. Iqbal, M. Rashid, and H. A. Habib, "A novel image retrieval based on visual words integration of sift and surf," *PloS One*, vol. 11, no. 6, p. e0157428, 2016.
[44] G. H. Liu, L. Zhang, Y. K. Hou, Z. Y. Li, and J. Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognition*, vol. 43, no. 7, pp. 2380-2389, 2010.
[45] P. S. Rao, E. Vamsidhar, G. Raju, R. Satapati, and K. Varma, "An approach for cbir system through multi layer neural network," *International Journal of Engineering Science and Technology*, vol. 2, no. 4, pp. 559-563, 2010.
[46] C. H. Lin, R. T. Chen, and Y. K. Chan, "A smart content-based image retrieval system based on color and texture feature," *Image and Vision Computing*, vol. 27, no. 6, pp. 658-665, 2009.
[47] A. Krizhevsky and G. Hinton. (2009). Learning multiple layers of features from tiny images. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.9220&rep=rep1&type=pdf
[48] T. S. Chua, J. Tang, R. Hong, H. Li, Z. Luo, and Y. Zheng, "Nus-wide: A real-world web image database from national university of singapore," in *Proc. the ACM International Conference on Image and Video Retrieval*, 2009, pp. 1-9.
[49] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, 1998.
[50] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," in *Proc. NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.
[51] C. C. Lai and Y. C. Chen, "A user-oriented image retrieval system based on interactive genetic algorithm," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 10, pp. 3318-3325, 2011.
[52] W. Bian and D. Tao, "Biased discriminant euclidean embedding for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 545-554, 2009.
[53] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, Springer, 2019.
[54] M. Abadi, *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th Symposium on Operating Systems Design and Implementation*, 2016, pp. 265-283.
[55] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.
[56] F. Chollet, *et al.* (2015). Keras: Deep learning library for theano and tensorflow. [Online]. Available: https://www.datasciencecentral.com/keras-deep-learning-library-for-theano-and-tensorflow/
[57] D. Tao. (2009). Corel 1000 and corel 10000 image database. [Online]. Available: http://wang.ist.psu.edu/docs/related.shtml
[58] M. K. Alsmadi, "Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 30, no. 3, pp. 373-381, 2018.
[59] K. V. Madhavi, R. Tamilkodi, and K. J. Sudha, "An innovative method for retrieving relevant images by getting the top-ranked images first using interactive genetic algorithm," *Procedia Computer Science*, vol. 79, pp. 254-261, 2016.
[60] S. M. Youssef, "Ictedct-cbir: Integrating curvelet transform with enhanced dominant colors extraction and texture analysis for efficient content-based image retrieval," *Computers & Electrical Engineering*, vol. 38, no. 5, pp. 1358-1376, 2012.
[61] M. B. Rao, B. P. Rao, and A. Govardhan, "Ctdcirs: Content based image retrieval system based on dominant color and texture features," *International Journal of Computer Applications*, vol. 18, no. 6, pp. 40-46, 2011.

Md. Mohsin Kabir was born in Dhaka, Bangladesh. He received a B.Sc. degree in Computer Science and Engineering from the Bangladesh University of Business and Technology (BUBT). He is currently working as a Lecturer and a Research Assistant with the Department of CSE, BUBT. He also works as a Researcher with the Advanced Machine Learning Laboratory. He is always willing to learn new things with full enthusiasm and passion. He has experience working in Python, Keras, TensorFlow, Sklearn, and Scipy. His research interests include deep learning, image processing, computer vision, and natural language processing.

Adit Ishraq is currently a student in Computer Science & Engineering, Bangladesh University of Business and Technology. He is also working as a Researcher with the Advanced Machine Learning Laboratory. His research focuses on deep learning algorithms and computer vision. He is proficient in prominent frameworks, including TensorFlow, Keras, PyTorch, NumPy, Matplotlib, etc.

Kamruddin Nur (Senior Member, IEEE) is currently serving as an associate professor in the Department of Computer Science at American International University-Bangladesh (AIUB). He also served as the chairman in the Department of Computer Science and Engineering at Stamford University Bangladesh (SUB) and Bangladesh University of Business and Technology (BUBT). Dr. Nur completed his PhD from UPF, Barcelona, Spain, Masters from UIU, and Bachelor from Victoria University of Wellington (VUW), New Zealand. Dr. Nur authored many prestigious journals and conferences in IEEE and ACM, served as TPC members, and reviewed articles in IEEE, ACM, Springer journals, and conferences. His research area includes Ubiquitous Computing, Computer Vision, Machine Learning, and Robotic Automation.

M. F. Mridha (Senior Member, IEEE) received the Ph.D. degree in AI/ML from Jahangirnagar University, in 2017. He joined the Department of Computer Science and Engineering, Stamford University Bangladesh, in June 2007, as a Lecturer, where he was promoted a Senior Lecturer and an Assistant Professor, in October 2010 and October 2011, respectively. Then, he joined UAP, in May 2012, as an Assistant Professor. He is currently working as an associate professor with the Department of Computer Science and Engineering, Bangladesh University of Business and Technology. He also worked as a faculty member of the CSE Department, University of Asia Pacific, and as a graduate coordinator, from 2012 to 2019. His research experience, within both academia and industry, has resulted in over 80 journal and conference publications. His research interests include artificial intelligence (AI), machine learning, deep learning, big data analysis, and natural language processing (NLP). For more than ten years, he has been with the master's and undergraduate students, as a supervisor of their thesis work. He has served as a program committee member of several international conferences and workshops. He also served as an associate editor in several journals.