

Article

A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion

Shahbaz Sikandar ¹, Rabbia Mahum ¹ and AbdulMalik Alsalmán ^{2,*}

¹ Computer Science Department, University of Engineering and Technology Taxila, Taxila 47050, Pakistan

² Computer Science Department, King Saud University, Riyadh 11451, Saudi Arabia

* Correspondence: salman@ksu.edu.sa

Abstract: The multimedia content generated by devices and image processing techniques requires high computation costs to retrieve images similar to the user's query from the database. An annotation-based traditional system of image retrieval is not coherent because pixel-wise matching of images brings significant variations in terms of pattern, storage, and angle. The Content-Based Image Retrieval (CBIR) method is more commonly used in these cases. CBIR efficiently quantifies the likeness between the database images and the query image. CBIR collects images identical to the query image from a huge database and extracts more useful features from the image provided as a query image. Then, it relates and matches these features with the database images' features and retakes them with similar features. In this study, we introduce a novel hybrid deep learning and machine learning-based CBIR system that uses a transfer learning technique and is implemented using two pre-trained deep learning models, ResNet50 and VGG16, and one machine learning model, KNN. We use the transfer learning technique to obtain the features from the images by using these two deep learning (DL) models. The image similarity is calculated using the machine learning (ML) model KNN and Euclidean distance. We build a web interface to show the result of similar images, and the Precision is used as the performance measure of the model that achieved 100%. Our proposed system outperforms other CBIR systems and can be used in many applications that need CBIR, such as digital libraries, historical research, fingerprint identification, and crime prevention.



Citation: Sikandar, S.; Mahum, R.; Alsalmán, A. A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion. *Appl. Sci.* **2023**, *13*, 4581. <https://doi.org/10.3390/app13074581>

Academic Editors: Jiangyun Li and Tianxiang Zhang

Received: 23 February 2023

Revised: 10 March 2023

Accepted: 31 March 2023

Published: 4 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, images are widely used, and the amount of graphics, photos, and images being generated is ever-increasing. This large amount of data needs data outsourcing services such as cloud storage and computing. Additionally, it requires content-based search and retrieval solutions [1]. The data have been increased excessively, yet content-based image retrieving is still a tedious job [2]. Retrieval of images is essential because images are used for different applications, such as crime prevention, biodiversity, information system, historical research, fingerprint identification, and medicine [3]. Generally, three image retrieval techniques are used: semantic-, text-, and content-based [4]. With the rapidly increasing requirement of CBIR in digital libraries, military, education, and architectural design, CBIR has become a more active research topic in recent years [5]. Retrieval of a huge amount of data is the main problem in traditional database technology, as the requirements of the image database are not filled by the traditional text object database [6]. The traditional method was based on textual annotation for image retrieval; images are interpreted with text and then recaptured from the database that focuses on text instead of visual features for image search [7].

Traditional image retrieval techniques are not suitable in terms of time and space, which encourages new technology to be developed [8]. Most traditional methods add

Metadata for retrieving images, such as descriptions, keywords, and captions [9]. Textual annotation of images relies on manual labor to associate the keywords with images, which cannot describe the ambiguity of image content and its diversity [10]. CBIR (Content-Based Image Retrieval) systems aim to improve the accuracy of image retrieval by focusing on prominent features of the images. By analyzing the features of an image, CBIR systems can provide more accurate and relevant results for a given search query. However, most of the methods of CBIR used only one to two feature descriptors that provide results with low accuracies, such as the scale-invariant feature transform (SIFT) descriptor, which can only reliably recognize objects when there is clutter. However, low illumination and poorly located key points inside an image caused SIFT descriptor to fail and lower the performance of CBIR. In this paper, pre-trained models of deep learning are used to extract the most prominent features. CBIR is the search that focuses on the image content instead of the Metadata, such as the description associated with the image, keyword, and tags [11]. CBIR was introduced in the 1990s to minimize the problem associated with the laborious text-based image retrieval system [12]. CBIR helps to extract or retake similar image content from a huge database and multimedia applications. It considerably reduces image retrieval time compared to unstructured and tedious browsing [13]. The visual content of the images is retrieved from the database and then stored as feature vectors [14]. The CBIR method calculates the feature vectors of the database and characterizes the visual properties against them. These image signatures or feature vectors act as feature databases and are used for similarity measures between the images of the repository and the query image [15]. CBIR shows the potential usefulness in the past twenty years of development in different real-world applications [16]. Image contents include image structure, shape, contrast, and color. The massive growth of visual information requires a capable system that retrieves images and video libraries from huge databases [7].

Many image retrieval methods exist, such as in [17]; authors used deep auto encoder for CBIR. Although multiple methods of CBIR exist, the semantic gap issue is one of the big problems between the high-level semantics that humans observed and the low-level pixels of images that machines captured [18]. DL algorithms are used in multiple diverse fields, such as in object detection present in the surveillance video [19], leaf disease detection [20], video summarization [21], and detection of glaucoma in fundus imaging [22], as well as in CBIR. Nowadays, CBIR is the most commonly adopted system for retrieving images based on feature extraction. The feature extraction process is the most prominent step in the CBIR method [23]. Therefore, in this paper, we extract the feature with the help of two different DL models and fuse them to get the most prominent features. The following are the key contributions:

- We proposed a novel system to capture the top five most identical images among 10,972 images, focusing on their visual features;
- The algorithm synthesizes two deep learning models: VGG16 and Resnet50, and KNN machine learning. We have extracted the features of images using Resnet50 and VGG16 and fused them to get better results and the most prominent features;
- To increase the performance of the proposed CBIR method, images are pre-processed, i.e., all the images are initially transformed from RGB to BGR for feature extraction, and then each color channel is zero-centered to the ImageNet dataset without scaling;
- We built a web interface to show the result or performance of the proposed CBIR system;
- We used the images from three different datasets: the coral-images dataset, the natural images dataset, and the fruits 360 dataset [24–26]. The images of these three datasets are used to retrieve similar images out of 28 different classes.

The remaining sections are as follows. Section 2 describes the related work. Section 3 concentrates on the CBIR system's methodology and implementation. Section 4 shows the results of our proposed CBIR system. Section 5 compares the proposed CBIR system with the existing ones. Finally, Section 6 concludes the work.

2. Related Work

CBIR is based mainly on three techniques: semantic, text, and content [4]. With the rapidly increasing requirement of CBIR in digital libraries, military, education, and architectural design, CBIR has become a more active research topic in recent years [5]. In [27], Zahid Mehmood proposed a technique for retrieving images that focuses on bags of visual words with the triangle histogram's weighted average. The proposed method uses bags of word models and spatial image contents. This approach reduces the overfitting problem on a dictionary of larger size. SIFT feature vector was used and then the k-means clustering technique was used so that the nearest features are clustered or Grouped. The authors used the Hellinger kernel function to calculate how similar the query picture and database images are. SVM with the Hellinger kernel was applied due to its simplicity in classification. To evaluate the scaling factor, the SVM Hellinger kernel function has been fixed using 10-fold cross-validation. Although the visual words are depicted in a histogram without considering where they are in the 2D image space, the primary issue is the lack of spatial information in BoVW-based image representation [28]. The authors in [29] proposed a CBIR method that used color, texture, shape, and grey-level co-occurrence matrix (GLCM) featuring contrast, correlation, homogeneity, and energy of images to retrieve similar pictures from the database. The Corel dataset was used to extract similar images and retrieve them using the meta-heuristic algorithm.

R. Rani Saritha [30] proposed a technique for retrieving images by combining the features of edge directions, texture features, color histogram, edge histogram, etc., using the deep learning model CNN. In the proposed method, content-based images were extracted from the image database. After a series of operations, such as selection removal edge directions, texture features, color histogram, and edge histogram, features were obtained and saved as small signature files. Signatures were matched to the content-based signature because similar images have similar signatures. Weights were used to standardize the distance coefficients after measuring the distances between the various characteristics. The distance values were used to save and index these normalized coefficients. The proposed model provides only 96% accuracy on the dataset of more than 1000 images. However, this accuracy can be improved by using some image pre-processing techniques. The aim of [31] was to introduce a method for image retrieval constructed on YCbCr color with a canny edge histogram and discrete wavelet transform. This method increases the performance of the content-based image retrieval framework. To identify image edges, Rehan Ashraf presented a CBIR approach based on color characteristics and the YCbCr color space [31]. To create a final RGB image containing edge and histogram information, the edge features were obtained using the Canny-edge feature descriptor and merged with the Y-luminance, Cr, and invariant Cb matrices. The discrete wavelet transforms served as the final feature vector, and the Manhattan distance was used to compare the query image to database images. The proposed system's accuracy was 73%.

Uzma Sharif et al. [32] proposed a system that focuses on the fusion of two feature descriptors: binary robust invariant scalable key points (BRISK) and scale-invariant feature transform (SIFT). Due to the SIFT descriptor's varying illumination, rotation, noise, and invariance to scale, it can only reliably recognize objects when there is clutter. However, low illumination and poorly located key points inside an image caused SIFT descriptor to fail and lower the performance of CBIR. In classification-based applications, the BRISK descriptor offers high-quality and flexible performance. When applied to poorly localized key points in an image along an object's edges, the BRISK descriptor outperforms the SIFT descriptor [32]. However, the proposed system achieved only 39.1% Mean Average Precision (MAP). Niranjana Sampathila1 et al. [33] proposed a method for CBIR from the MRI images by using features of images such as texture, shape, and color. The proposed method retrieves and represents the brain MRI images from a vast database that is closely similar to that query image. The proposed methodology was used for selecting distinct planes (sagittal, transverse, and coronal) in brain MRI images from a dataset of normal and demented people who were assessed. The features were determined using Haralik's

features based on Grey level co-occurrence and histogram-based cumulative distribution function (CDF). The K-Nearest Neighbor algorithm was used for the image retrieval technique to determine the minimum distance or similarity between the query image and the database images. The performance Matrices that were calculated for the model were Precision and Recall. The proposed model obtained 95.5% average accuracy. However, it can be increased by using the DL models for the feature extraction process.

In [34], the authors introduced a method that supports cloud server-based local feature extraction such as SIFT, image similarity scoring, and Index building. They proposed a multi-index for SIFT descriptors, the encrypted image was loaded, and the model returned similar images from the database in response to a query image. Exact Euclidean LSH was applied on SIFT descriptors, and secure block identifier computational protocol was used as the building block of image retrieval and multi-index construction. Due to the SIFT descriptor's varying illumination, rotation, noise, and invariance to scale, it can only recognize objects reliably when there is clutter. However, low illumination and poorly located key points inside an image caused SIFT descriptor to fail and lower the performance of CBIR. The purpose of the [35] study was to propose an approach containing the complete structural information about an image extracted from the image by descriptors such as local binary patterns and GLCM. GLCM descriptor, which was used to extract the dominant local binary pattern (LBP) and the statistical traits for classifying textures in images. To reduce computing load during the classification process, the proposed method employs particle swarm optimization (PSO) to reduce the number of features. Support vector machine, decision tree, and K-nearest neighbor, the three classifiers, were trained, and Euclidean distance was employed as a similarity metric. SVM performs the best among these models and provides 85% precision; however, it can be increased.

Muhammad Kashif et al., in their [36] study, proposed a new framework for the efficient retrieval of lung disease for a patient with the combination of different descriptors, such as local phase quantization (LPQ), local ternary pattern (LTP), and discrete wavelet transform. Joint mutual information (JMI)-based feature was also used to minimize duplication and select an ideal feature set for retrieving CIs. The similarity was measured to create a balanced graph by combining semantic and visual information in an equal ratio. The final degree of similarity between the images was calculated using the shortest path. The computations were performed with distance metrics such as cosine, city block, correlation, Spearman, and Euclidean. However, the proposed approach is inefficient in terms of computational cost because manual feature descriptors were used, such as LPQ, LTQ, and DWT, to extract features from images that take more time and resources. Authors in [37] used color and edge descriptors for the extraction of features and Euclidean Distance, L1, L2, and cosine to check the similarity. Similarly, in [38], only the SURF feature descriptor was used to extract the visual features. A comparison of related work is listed in Table 1.

Table 1. Comparison of related work.

Ref No	Year	Features	Similarity Measure	Model Used
[27]	2018	SIFT	Hellinger kernel	SVM
[31]	2018	DWT	Manhattan distance	-
[37]	2018	HSV histogram, color, and edge descriptor	Cosine, L1, L2, Euclidean	-
[29]	2018	GLCM	-	Metaheuristic algorithm
[32]	2019	BRISK	-	-
[34]	2019	SIFT	Euclidean Distance	-
[30]	2019	Edge and color histogram, texture	-	CNN
[36]	2020	LPQ, LTQ	City block, cosine, Euclidean	-
[35]	2021	GLCM	Euclidean Distance	KNN
[33]	2022	GLCM	-	KNN
[39]	2022	GLCM	Manhattan distance	Alex Net
[40]	2022	Variational Auto-encoder	Cosine similarity	-
[41]	2022	DCNN	Euclidean distance	Nearest Neighbor

Table 1. Cont.

Ref No	Year	Features	Similarity Measure	Model Used
[42]	2022	High-Level features by using DCNN	-	K-means clustering
[43]	2023	SVD, UMLBP, Color autocorrelogram	Mahalanobis distamce	-
[38]	2023	SURF	-	-
[44]	2023	CNN	Hamming Distance	Random Forest

3. Materials and Methods

In this paper, we combine three datasets to form a huge dataset that consists of 10,972 images. This section will discuss the dataset used, the DL models used for feature extraction, and the machine learning model used for clustering the features. Initially, images are converted from RGB to BGR and then provided to the improved deep-learning models. After that, the feature extraction procedure uses CNN architectures, and the results are acquired. Finally, the results, obtained using the python environment, are presented in detail in Section 4.

3.1. Deep Learning

An essential component of artificial intelligence is deep learning. Deep learning uses numerous non-linear processing layers to extract features and their conversion. Different layers are sequentially ranked to form a deep network architecture. These layers take input from the previous layer's output and are classified as consecutive. Many different architectures are employed and preferred in DL; however, in this study, we examined the CNN architectures of DL. Deep learning models called CNNs are typically employed for categorization tasks [45].

Models already trained are injected into the model's development process as part of the transfer learning process. Therefore, pre-trained models are used rather than building a model from scratch, which reduces time and space complexity and even improves the model's accuracy. In this paper, two pre-trained models are used: ResNet50 and VGG16. Details of these two models are given in the next section.

3.2. Structure of System

This paper uses ResNet50 and the VGG16 architectures of deep learning. ResNet50 won the competition that was held in 2015 on ILSVRC ImageNet 2015 as the base [46]. Instead of starting CNN from scratch, the ResNet50 and VGG 16 models are used as the base transfer learning methods [47,48]. Thus, the accretion ResNet50 and VGG16 of a pre-trained model are utilized instead of creating a new model. We used ResNet50 because it successfully solves many problems related to image processing and classification of biomedical images [49]. Moreover, Resnet50 reduces computational costs and time complexity. We used VGG16 for the feature fusion purpose to achieve high accuracy using the final feature vector.

The input layers of both models are modified to accept information in the forms of 224, 224, and 3. After the input layer is modified and used as the foundation for the new model, the ResNet50 and VGG16 models no longer have any input, pooling, fully connected, softmax, or classification layers. The hybrid model adds a new GlobalMaxpooling2D layer to retrieve the single vector array. The architecture of the proposed CBIR is shown in Figure 1, and the parameters of the network are listed in Table 2.

Table 2. Parameters of the network's layers.

Layers	Output Shape	Parameters	Feature	Total Features
ResNet50 (Functional)	7, 7, 2048	23,587,712	-	
Global_max_pooling2d	2048	0	2048	
VGG16 (Functional)	7, 7, 512	14,714,688	-	2560
Global_max_pooling2d	512	0	512	

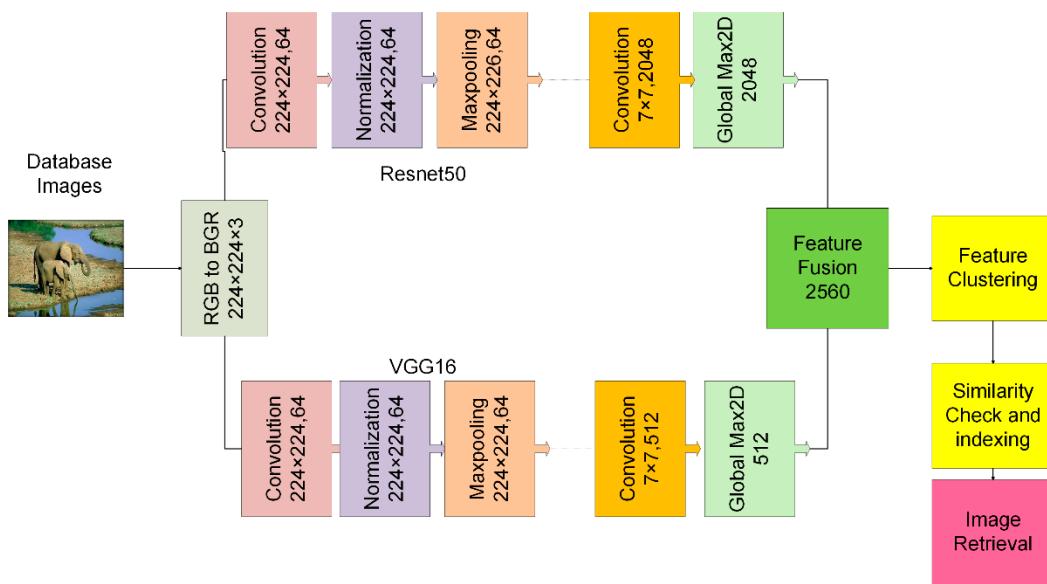


Figure 1. The proposed CBIR.

Database images are initially pre-processed and transformed from RGB to BGR and then passed through the two pre-trained deep learning models, ResNet50 and VGG16, for extracting features. GlobalMaxpooling2D layer is added to convert the feature vector into a one-dimensional array. A total of 2048 features are extricated with ResNet50 and 512 with VGG16. After the feature extraction, all the features are combined to create the final feature vector.

Similar features are clustered together by applying the machine learning model k Nearest Neighbor. K-NN is a technique utilized to retrieve images that closely resemble the query image. It works by identifying the nearest neighbors to the query image and retrieving their corresponding indices. Essentially, K-NN makes groups of similar images and retrieves those that are most similar to the query image. After clustering the features to determine the distance between similar images, the Euclidean distance similarity measure is used. In the image retrieval phase, the images comparable to the query image or those closest to it are indexed and retrieved. Figure 2 depicts the image retrieval process.

3.2.1. Input Layer

The input layer is the initial layer to supply input to the models. The high selection of the dimension of the input size of the image in this layer can extend the memory requirement during the training. Thus, we select the dimension of the image input layer to $224 \times 224 \times 3$.

3.2.2. Activation Function

The deep neural network's Relu layer is the activation layer. The Relu layer set the negative input value to zero to reduce the problem of overfitting. The network will run faster if it gets a zero value on the negative axis. In this paper, we use Relu as an activation function. The activation function of Relu is listed in Equation (1).

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (1)$$

The computational load of the Relu layer is less than the other activation functions. Therefore, Relu is preferred [50]. The operation of the Relu activation function is shown in Table 3.

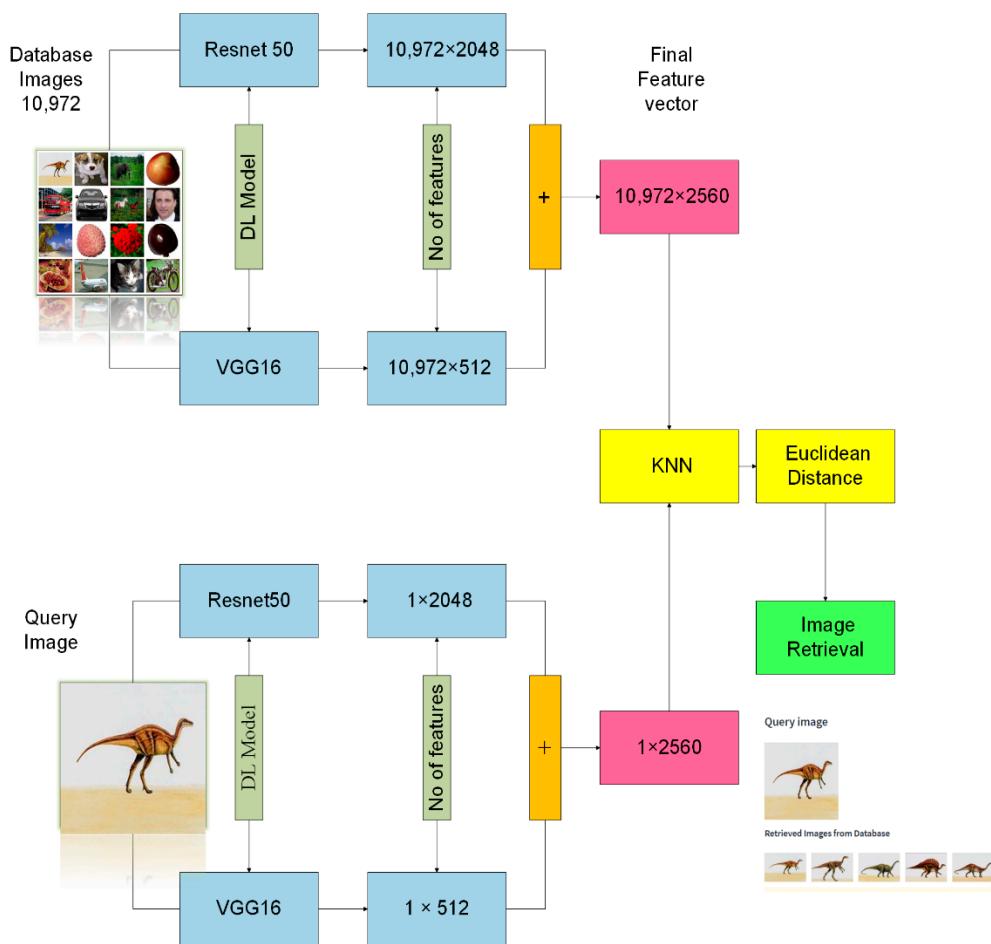


Figure 2. The process of image retrieval.

Table 3. Operation of Relu Activation Function.

t	$f(t)$	$F(t)$
-10	$f(-10)$	0
-150	$f(-150)$	0
100	$f(100)$	100
125	$f(125)$	125

The process of the Relu activation function on a 4×4 pixel image is shown in Figure 3.

3.2.3. Convolution Layer

The convolution layer serves as the foundation of the convolutional neural network. The transformation layer is another name for the convolution layer. A convolution filter is slid across the entire image during the convolutional process. The convolutional layer's objective is to create a property map or feature map in the convolution layer. A set of linear filters are used in this layer's convolution process, as illustrated in Equation (2).

$$(O_f)_{ij} = (W_f * R)_{ij} + b_{ij}, \quad (2)$$

where (i, j) are the index of pixel point, f is the feature map index, b is the bias and W is weight, R is input data, and $(O_f)_{ij}$ is the feature map's output value.

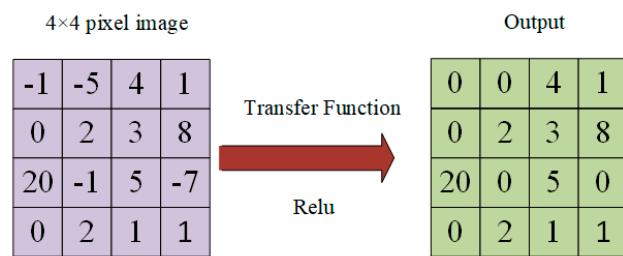


Figure 3. The process of ReLU activation function.

3.2.4. Normalization Layer

The normalization step improves the network's efficiency in DL models. In addition to improving the model's performance, normalizing the data dimensions allows the input data to be plotted and scaled to a certain range. Batch normalization emphasizes the standardizing of the inputs to any particular layer. Standardizing the inputs means the network should have approximately unit variance and zero means to any layer in the network. Batch normalization transforms the input by dividing the subtraction of the input mean in the current mini-batch with the standard deviation in the current mini-batch. Equations (3) and (4) illustrate the normalization process.

$$\hat{x}^k = x^k - E(x^k) / \sqrt{\text{var}(x^k)} \quad (3)$$

Here, \hat{x}^k is the normalized value of the k^{th} hidden neuron and $E(x^k)$ is the dimension's average. The variance of the k^{th} hidden unit is represented here by $\sqrt{\text{var}(x^k)}$, and β and γ are the two learnable variables.

$$y^{(k)} = \gamma^k x^k + \beta^k, \quad (4)$$

3.2.5. Dropout Layer

Deep learning models perform well when massive data are available for network training. As a result, memorization can take place as the network is being trained. However, memorization of the network creates the problem of network overfitting, which is undesirable. The memorization problem occurs when there are deeper hidden layers in the DNN. To avoid network memorization, some nodes must be disabled. When some nodes are disabled, the memorization process is halted. Hence, it improves the network performance by applying a dropout layer in the network [51]. An example of a dropout operation in a network is shown in Figure 4.

3.2.6. GlobalMaxpooling 2D Layer

Globalmaxpooling2D layer performs a similar operation as the 2D max pooling. The block's whole input size serves as the globalmaxpooling2d layer's pool size. globalmaxpooling2d layer returns a single max value for each input channel. It is used for down-sampling. It down-samples the input value for each input channel along its spatial dimensions by taking the highest value over the input window.

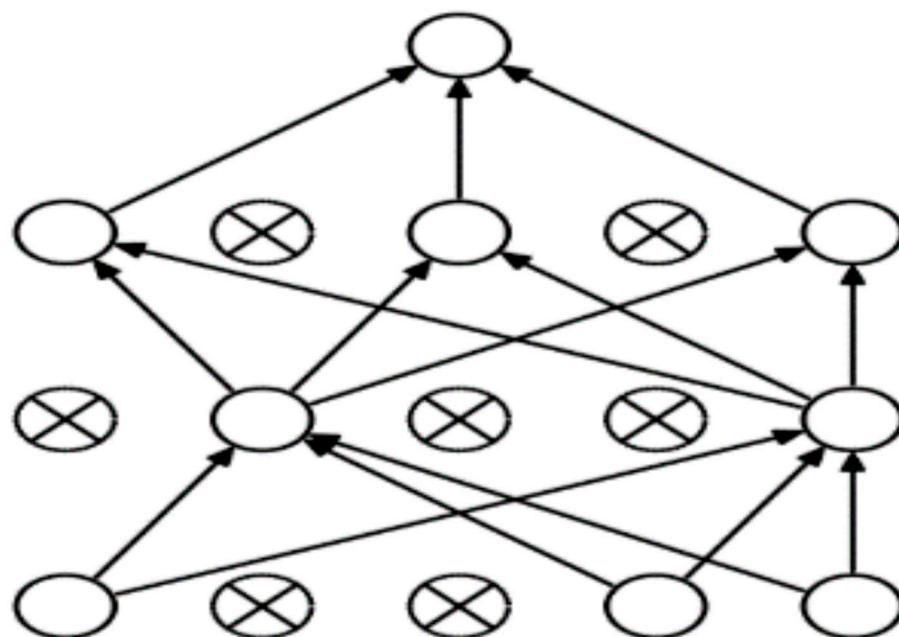


Figure 4. Dropout process [52].

4. Experimentation

4.1. Dataset Used

We used a combination of three datasets collected from the Kaggle website. There are 10,972 images of 28 different classes, such as elephants, horses, airplanes, beaches, buses, cars, etc. Training and test folders are available for all classes. Images of the training folder are used for feature extraction, while images of the test folder are used as the query image. Some used image samples are presented in Figure 5, and the dataset is presented below.

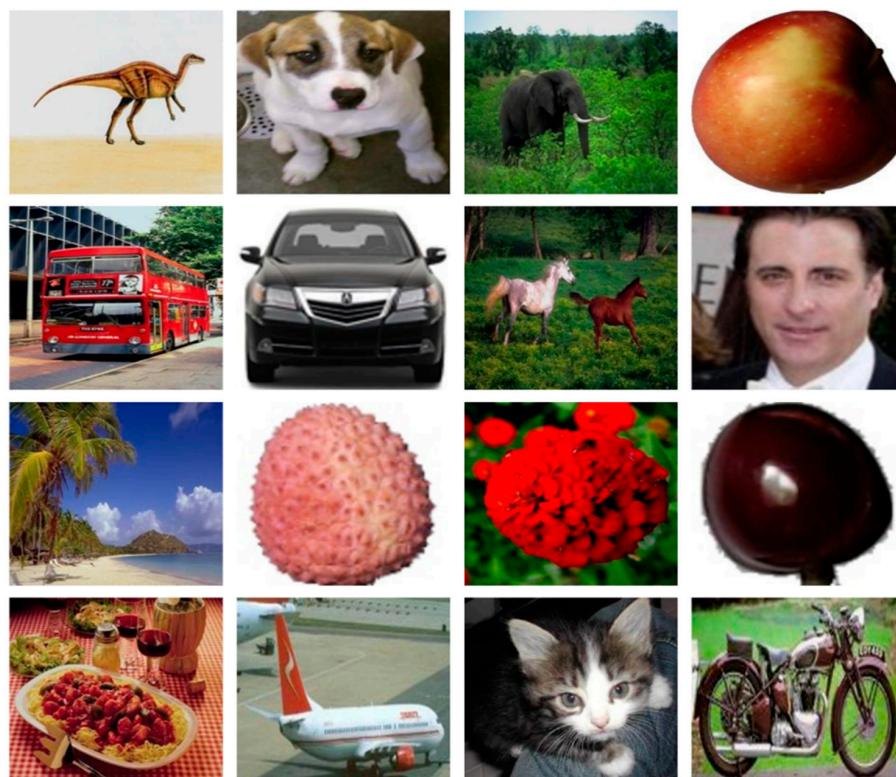


Figure 5. Samples of images used in the dataset.

4.1.1. Corel-Images

We used the nine classes of Corel-images dataset that is publicly available on the Kaggle website. The classes we used from this dataset are beaches, dinosaurs, flowers, food, horses, mountains, buses, elephants, and snow. There are 810 images, 90 in each class, out of 10,972 images taken from the Corel-image dataset.

4.1.2. Natural-Images

The natural-images dataset consists of 6899 images and comprises eight classes: airplanes, flowers, persons, cars, motorbikes, cats, fruit, and dogs. There are 1000 images of fruits, 727 images are airplanes, 968 images are relevant to cars, 885 images are cats, 702 images belong to the dog category, 788 to motorbikes, and 843 images belong to flower class reaming, and 986 images belong to the person class.

4.1.3. Fruits 360-Images

Fruits 360 is a huge dataset with 90,483 images; however, we used 3233 images from this dataset. We collect 3233 images from this dataset with 11 classes. Nine classes belong to different categories of a class apple, and two classes belong to the pear class.

4.2. Evaluation and Performance

The CBIR system's effectiveness can be evaluated using Precision, as shown in Equation (5).

$$\text{Precision} = \frac{\text{Number of suitable retrieved images}}{\text{Total number of retrieved images}}, \quad (5)$$

Precision is the percentage of accurately retrieved images by the proposed model. The performance of our proposed system in CBIR is shown in Figure 6.

The Precision percentages of our proposed system on coral images, natural images, and the fruits 360 dataset are listed in Table 4.

Table 4. The Precision of the proposed CBIR.

Matrices Used	Corel-Images	Natural-Images	Fruits 360-Images	Overall
% Precision	100	100	100	100

As in Figure 6, we retrieved five images, and all five images of each class are identical to the query image, so the precision of our model is 100% on coral-image, natural-image, and fruits 360 dataset images. We evaluated the suggested CBIR method by conducting tests on various datasets, such as Caltech 101 and WANG. The results of our experiments showed remarkable performance in specific categories such as airplanes, motorbikes, cars, and elephants. Moreover, the model exhibited high precision in identifying images belonging to the beach class of the WANG dataset. In addition, we assessed the proposed method's effectiveness by conducting tests on randomly obtained images from Google. Our experimental results demonstrated that the suggested CBIR method surpasses existing methods in terms of accuracy and speed, and the proposed system exhibits considerable retrieval capabilities. In Figure 7, the performance of the proposed model on randomly obtained images from Google is presented.

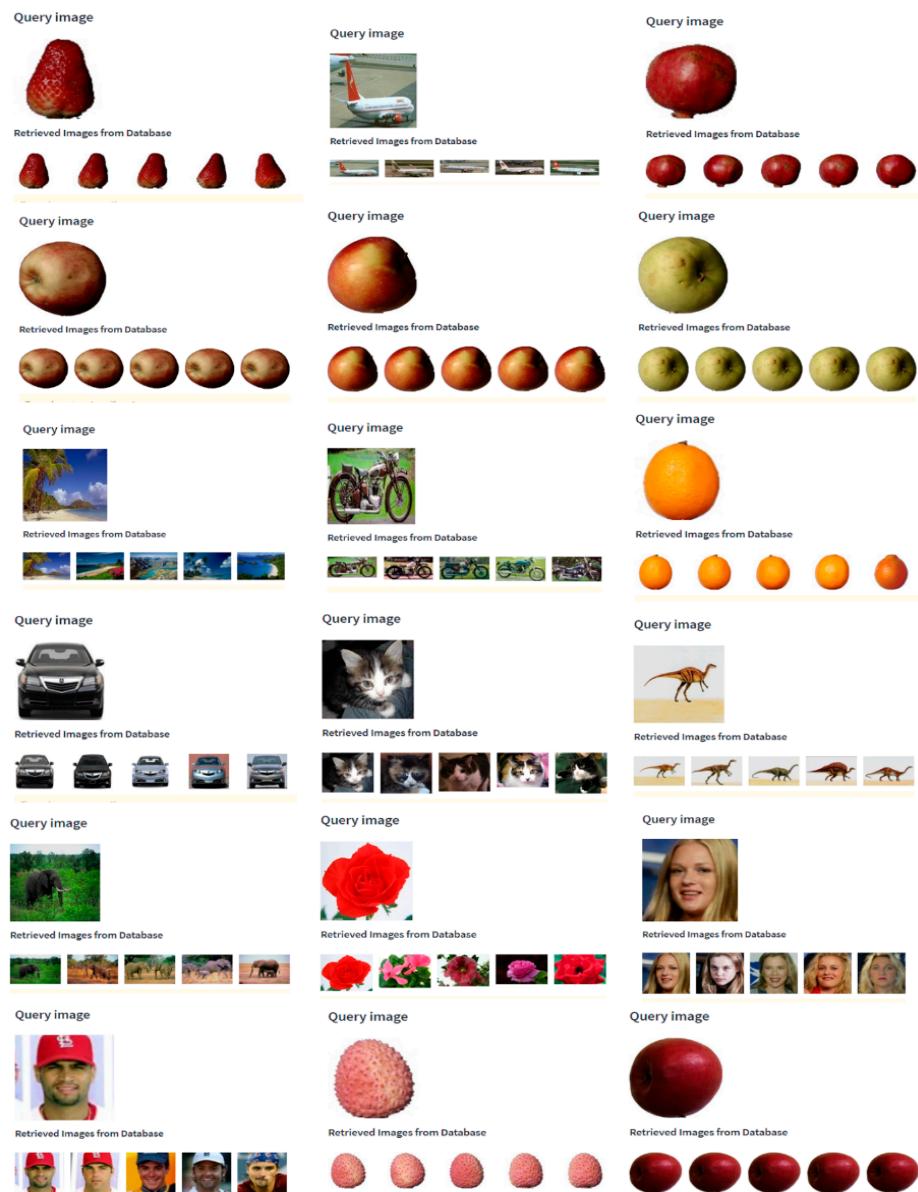


Figure 6. Samples results of CBIR.

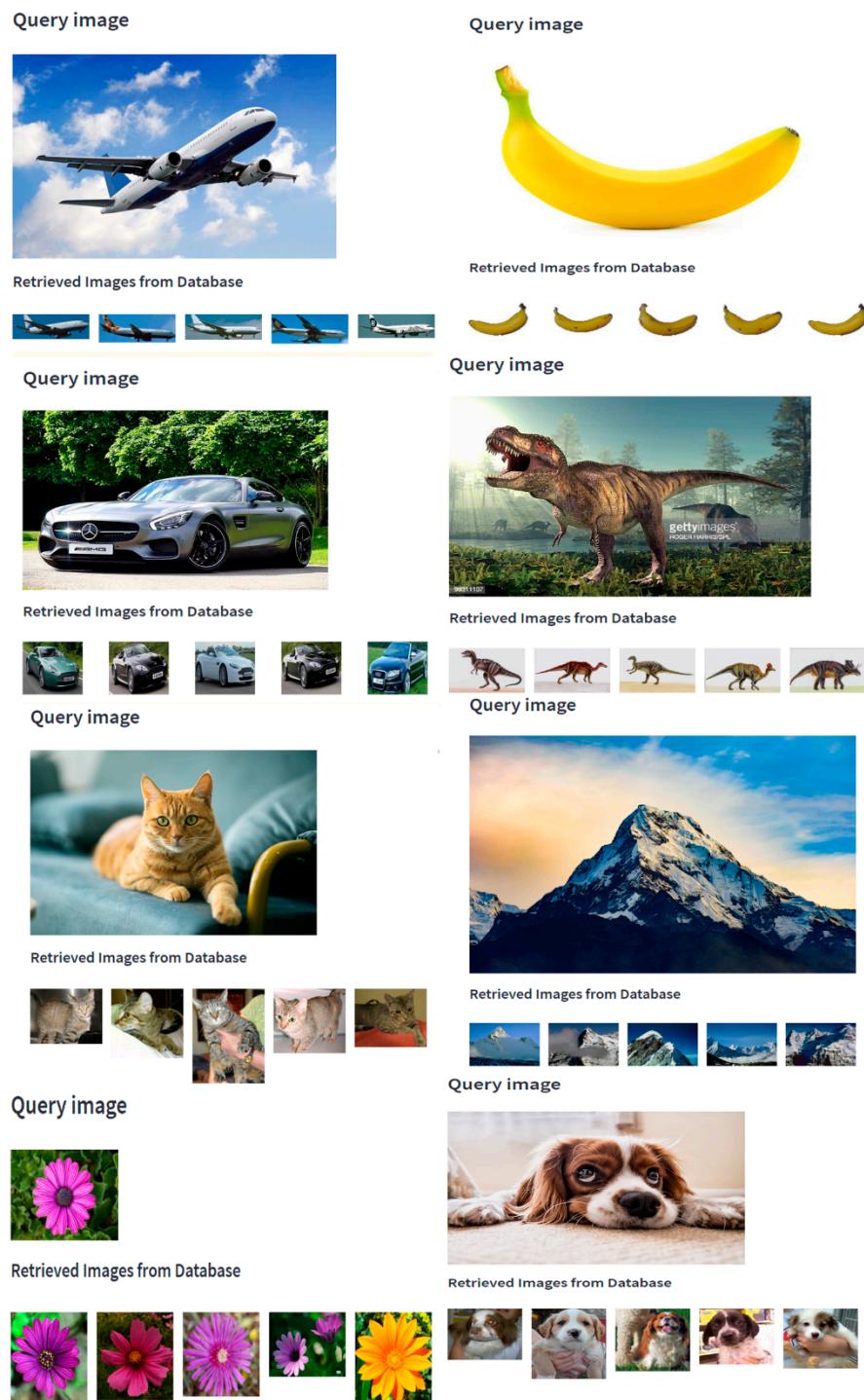


Figure 7. Performance of CBIR on random images.

5. Comparison with Existing Techniques

The goal of the [53] study was to suggest an evolutionary method for CBIR that uses k nearest neighbors to choose the best possible solution of the diffusion parameters. To obtain a better set of parameters, genetic algorithms were used for this optimization problem. The proposed system achieved 97.32% precision on the “oxford5k”, “Paris 5k”, and “flickr 1M” datasets. Tao HE [54] proposed a method for the CBIR from the “coil20” and “coral1K” datasets using SIFT feature extractor. SIFT feature extractor results were compared with the other two feature extractors, HOG and SURF. SIFT performs well in retrieving similar images compared to the other two feature descriptors, HOG and SURF. The proposed

model achieved 96% overall precision. The authors of [55] used three datasets, “Corel 1k”, “Corel 5K”, and Caltech 256, for the extraction of a similar image in the domain of content-based image retrieval. The proposed methodology provides 99.4% precision on the Corel1K dataset. In this [38] study, the SURF feature descriptor was used for content-based image retrieval. The proposed method provides 86.06% precision on the retrial of the top 10 images from the “corel1K” dataset.

The purpose of the [37] study was to retrieve images from 10 different classes, such as buildings, Africa, elephants, flowers, beaches, horses, buses, dinosaurs, food, and mountains from the “Corel 1K” dataset. The entire collection of images was used to extract features, which were then stored as feature vectors. Features were extracted using different feature descriptors, such as Color and Edge Directivity Descriptors (CEDD), HSV histogram, Color moments, and Binarized Statistical Image Features (BSIF). The feature retrieved with the HSV histogram’s help involves quantization, color space conversion, and histogram computation. Following the feature extraction, a similarity assessment was performed, comparing the query image to other database images to identify the images that were similar to it. The feature vector was constructed using a variety of distance metrics, including the Mahalanobis, City block, Minkowski distance, Euclidean, Chebyshev, cosine, and L1 and L2 distances, to determine how similar the query image and the other images were. The proposed model has a precision rate of 87.5%. Similar to this [56] study, a novel CNN with a relevant feedback method was proposed for CBIR. The proposed method was tested on three datasets, “OT”, “Caltech 101”, and “Corel 1K”. The proposed CNN performs best on Caltech 101 dataset and provides 98.78% precision.

The authors of [57] proposed a bi-layer system for CBIR divided into two modules: the first module was designed to extract the features in terms of color, texture, and shape from the images, and the second module has two layers. To extract the form, texture, and indices of the M most similar images to the query image, all images in the “Corel 1K” dataset were compared to it. In the second layer, M images are compared with the query image for shape and color feature space retrieved from the previous layer. Finally, F images that are more similar to the query image are obtained as an output. The overall average precision of the proposed module was 92.2%. The authors of [58] used deep learning models such as VGG19, Inception v3, dense net, mobile net, and Inception ResNet for the CBIR Image DB2000, Caltech101, and Corel 1K dataset. Inception ResNet provided a higher precision rate of 96.12% among all these models. Figure 8 depicts a comparison between the proposed system and current methods, and Table 5 lists the comparison in terms of precision.

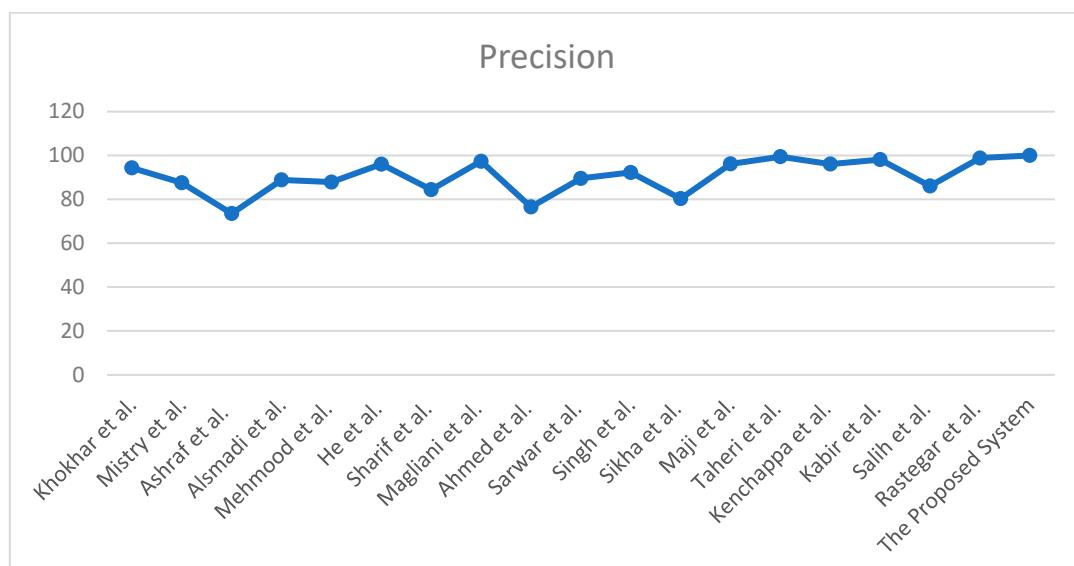


Figure 8. Comparison with existing techniques.

Table 5. Comparison of precision with existing techniques.

Ref	Year	Dataset Used	%Precision
Khokhar et al. [59]	2017	WANG	94.3
Mistry et al. [37]	2018	WANG	87.5
Ashraf et al. [31]	2018	WANG	73.5
Alsmadi et al. [29]	2018	Corel	88.83
Mehmood et al. [27]	2018	Corel	87.85
He et al. [54]	2018	Coil20, corel1k	96
Sharif et al. [32]	2019	Corel, Caltech 256	84.39
Magliani et al. [53]	2019	Oxford5k, paris5K, flicker1M	97.32
Ahmed et al. [60]	2019	Caltech101, corel1K	76.5
Sarwar et al. [61]	2019	WANG, Holidays	89.5
Singh et al. [57]	2020	Corel1K	92.2
Sikha et al. [62]	2021	WANG	80.3
Maji et al. [58]	2021	Imagedb2000, Caltech101, corel1K	96.12
Taheri et al. [55]	2022	Corel1K, Corel5K, caltech256	99.4
Kenchappa et al. [39]	2022	Corel, VOC	96.03
Kabir et al. [42]	2022	Corel10K, CAFIR10	98.1
Salih et al. [38]	2023	Corel1K	86.06
Rastegar et al. [56]	2023	OT, caltech101, corel1K	98.78
The Proposed System	-	Corel, Natural, fruits 360	100

6. Conclusions

This study proves that, by using pre-trained DL models, better features can be extracted. This can provide a better result in terms of precision concerning the manual feature descriptors derived by traditional methods such as SIFT, GLCM, wavelet, LBP, LTQ, etc. This work proposes a novel CBIR technique based on transfer learning to retrieve images from the massive database using DL and ML models. The image retrieval method of the proposed system consists of two modules. Feature extraction from the images is performed in the first module using pre-trained models ResNet50 and VGG16 to lower the computational cost. In the second module, features from the query image and features retrieved from database images are compared, and the most similar images are presented in response.

The effectiveness of the suggested CBIR approach is assessed on images from the natural-image dataset, fruits 360 dataset, and coral images using the evaluation metric precision. The proposed model was evaluated on the Caltech 101 dataset and exhibited exceptional performance in certain classes, such as airplanes, motorbikes, cars, and elephants. Additionally, the model demonstrated high precision on the beach class of the WANG dataset. Experimental findings show that, in comparison to existing methods, the suggested CBIR method is more accurate and quick, and the suggested system has significant retrieving capabilities. The suggested technique works better than the image retrieval system and achieves a 100% precision rate on a huge database with more than 10,000 images. However, one limitation of the features extracted using pre-trained deep learning models is that the system's retrieval outcomes may not be good if we query similar images from the database with images at various orientation angles. A rotation-invariant CBIR system can be introduced in the future, which can be the next improvement step.

Author Contributions: Methodology, S.S.; Validation, R.M.; Investigation, R.M. and A.A.; Data curation, R.M.; Writing—original draft, S.S.; Writing—review & editing, A.A.; Supervision, R.M. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was supported by a grant from the “Research Center of College of Computer and Information Sciences”, Deanship of Scientific Research, King Saud University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing does not apply to this article as authors have used publicly available datasets, whose details are included in the “experimental results and discussions” section of this article. Please contact the authors for further requests.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferreira, B.; Rodrigues, J.; Leitao, J.; Domingos, H. Privacy-preserving content-based image retrieval in the cloud. In Proceedings of the 2015 IEEE 34th Symposium on Reliable Distributed Systems (SRDS), Montreal, QC, Canada, 28 September–1 October 2015.
2. Dahake, P.A.; Thakare, S.S. Content based image retrieval: A review. *Int. Res. J. Eng. Technol.* **2018**, *5*, 1059–1061.
3. Pattanaik, S.; Bhalke, D. Beginners to content-based image retrieval. *Int. J. Sci. Eng. Technol. Res.* **2012**, *1*, 40–44.
4. Liu, G.-H.; Yang, J.-Y. Content-based image retrieval using color difference histogram. *Pattern Recognit.* **2013**, *46*, 188–198. [[CrossRef](#)]
5. Tian, D. Support Vector Machine for Content-based Image Retrieval: A Comprehensive Overview. *J. Inf. Hiding Multim. Signal Process.* **2018**, *9*, 1464–1478.
6. Yue, J.; Li, Z.; Liu, L.; Fu, Z. Content-based image retrieval using color and texture fused features. *Math. Comput. Model.* **2011**, *54*, 1121–1127. [[CrossRef](#)]
7. Shete, D.S.; Chavan, M.S.; Kolhapur, K. Content based image retrieval. *Int. J. Emerg. Technol. Adv. Eng.* **2012**, *2*, 85–90.
8. Rehman, M.; Iqbal, M.; Sharif, M.; Raza, M. Content based image retrieval: Survey. *World Appl. Sci. J.* **2012**, *19*, 404–412.
9. Yildizer, E.; Balci, A.M.; Hassan, M.; Alhajj, R. Efficient content-based image retrieval using multiple support vector machines ensemble. *Expert Syst. Appl.* **2012**, *39*, 2385–2396. [[CrossRef](#)]
10. Wang, X.-Y.; Yang, H.-Y.; Li, D.-M. A new content-based image retrieval technique using color and texture information. *Comput. Electr. Eng.* **2013**, *39*, 746–761. [[CrossRef](#)]
11. Sharma, M.; Batra, A. Analysis of distance measures in content based image retrieval. *Glob. J. Comput. Sci. Technol.* **2014**, *14*, 11–16.
12. Tzelepi, M.; Tefas, A. Deep convolutional learning for content based image retrieval. *Neurocomputing* **2018**, *275*, 2467–2478. [[CrossRef](#)]
13. Akgül, C.B.; Rubin, D.L.; Napel, S.; Beaulieu, C.F.; Greenspan, H.; Acar, B. Content-based image retrieval in radiology: Current status and future directions. *J. Digit. Imaging* **2011**, *24*, 208–222. [[CrossRef](#)] [[PubMed](#)]
14. Khokher, A.; Talwar, R. Content-based image retrieval: Feature extraction techniques and applications. In Proceedings of the International Conference on Recent Advances and Future Trends in Information Technology (iRAFIT2012), Patiala, India, 21–23 March 2012.
15. Irtaza, A.; Jaffar, M.A.; Aleisa, E.; Choi, T.-S. Embedding neural networks for semantic association in content based image retrieval. *Multimed. Tools Appl.* **2014**, *72*, 1911–1931. [[CrossRef](#)]
16. Xia, Z.; Wang, X.; Zhang, L.; Qin, Z.; Sun, X.; Ren, K. A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing. *IEEE Trans. Inf. Secur.* **2016**, *11*, 2594–2608. [[CrossRef](#)]
17. Krizhevsky, A.; Hinton, G.E. Using very deep autoencoders for content-based image retrieval. In Proceedings of the ESANN 2011, 19th European Symposium on Artificial Neural Network, Bruges, Belgium, 27–29 April 2011.
18. Wan, J.; Wang, D.; Hoi, S.C.H.; Wu, P.; Zhu, J.; Zhang, Y.; Li, J. Deep learning for content-based image retrieval: A comprehensive study. In Proceedings of the 22nd ACM International Conference on Multimedia, Orlando, FL, USA, 3–7 November 2014.
19. Mahum, R.; Irtaza, A.M.A.; Masood, M.; Nawaz, M.; Nazir, T. Real-time Object Detection and Classification in Surveillance Videos using Hybrid Deep Learning Model. In Proceedings of the 6th Multi Disciplinary Student Research International Conference (MDSRIC), Wah, Pakistan, 30 November–1 December 2021.
20. Mahum, R.; Munir, H.; Mughal, Z.-U.; Awais, M.; Khan, F.S.; Saqlain, M.; Mahamad, S.; Tlili, I. A novel framework for potato leaf disease detection using an efficient deep learning model. *Hum. Ecol. Risk Assess. Int. J.* **2023**, *29*, 303–326. [[CrossRef](#)]
21. Mahum, R.; Irtaza, A.; Nawaz, M.; Nazir, T.; Masood, M.; Mehmood, A. A generic framework for Generation of Summarized Video Clips using Transfer Learning (SumVClip). In Proceedings of the 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), Karachi, Pakistan, 15–17 July 2021.
22. Mahum, R.; Rehman, S.U.; Okon, O.D.; Alabrah, A.; Meraj, T.; Rauf, H.T. A novel hybrid approach based on deep CNN to detect glaucoma using fundus imaging. *Electronics* **2021**, *11*, 26. [[CrossRef](#)]
23. Murala, S.; Maheshwari, R.P.; Balasubramanian, R. Local tetra patterns: A new feature descriptor for content-based image retrieval. *IEEE Trans. Image Process.* **2012**, *21*, 2874–2886. [[CrossRef](#)]
24. Bian, W.; Tao, D. Biased discriminant Euclidean embedding for content-based image retrieval. *IEEE Trans. Image Process.* **2009**, *19*, 545–554. [[CrossRef](#)]
25. Roy, P.; Ghosh, S.; Bhattacharya, S.; Pal, U. Effects of Degradations on Deep Neural Network Architectures. *arXiv* **2018**, arXiv:1807.10108.
26. Mureşan, H.; Oltean, M. Fruit recognition from images using deep learning. *arXiv* **2017**, arXiv:1712.00580. [[CrossRef](#)]
27. Mehmood, Z.; Mahmood, T.; Javid, M.A. Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine. *Appl. Intell.* **2018**, *48*, 166–181. [[CrossRef](#)]
28. Philbin, J.; Chum, O.; Isard, M.; Sivic, J.; Zisserman, A. Object retrieval with large vocabularies and fast spatial matching. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 17–22 June 2007.

29. Alsmadi, M.K. Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm. *J. King Saud Univ.-Comput. Inf. Sci.* **2018**, *30*, 373–381. [[CrossRef](#)]
30. Saritha, R.R.; Paul, V.; Kumar, P.G. Content based image retrieval using deep learning process. *Clust. Comput.* **2019**, *22*, 4187–4200. [[CrossRef](#)]
31. Ashraf, R.; Ahmed, M.; Jabbar, S.; Khalid, S.; Ahmad, A.; Din, S.; Jeon, G. Content based image retrieval by using color descriptor and discrete wavelet transform. *J. Med. Syst.* **2018**, *42*, 1–12. [[CrossRef](#)] [[PubMed](#)]
32. Sharif, U.; Mehmood, Z.; Mahmood, T.; Javid, M.A.; Rehman, A.; Saba, T. Scene analysis and search using local features and support vector machine for effective content-based image retrieval. *Artif. Intell. Rev.* **2019**, *52*, 901–925. [[CrossRef](#)]
33. Sampathila, N.; Martis, R.J. Computational approach for content-based image retrieval of K-similar images from brain MR image database. *Expert Syst.* **2022**, *39*, e12652. [[CrossRef](#)]
34. Wang, Y.; Liu, F.; Pang, Z.; Hassan, A.; Lu, W. Privacy-preserving content-based image retrieval for mobile computing. *J. Inf. Secur. Appl.* **2019**, *49*, 102399. [[CrossRef](#)]
35. Garg, M.; Dhiman, G. A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants. *Neural Comput. Appl.* **2021**, *33*, 1311–1328. [[CrossRef](#)]
36. Kashif, M.; Raja, G.; Shaukat, F. An efficient content-based image retrieval system for the diagnosis of lung diseases. *J. Digit. Imaging* **2020**, *33*, 971–987. [[CrossRef](#)]
37. Mistry, Y.; Ingole, D.T.; Ingole, M.D. Content based image retrieval using hybrid features and various distance metric. *J. Electr. Syst. Inf. Technol.* **2018**, *5*, 874–888. [[CrossRef](#)]
38. Salih, S.F.; Abdulla, A.A. An effective bi-layer content-based image retrieval technique. *J. Supercomput.* **2023**, *79*, 2308–2331. [[CrossRef](#)]
39. Kenchappa, Y.D.; Kwadiki, K. Content-based image retrieval using integrated features and multi-subspace randomization and collaboration. *Int. J. Syst. Assur. Eng. Manag.* **2022**, *13*, 2540–2550. [[CrossRef](#)]
40. Lee, K.; Lee, Y.; Ko, H.-H.; Kang, M. A Study on the Channel Expansion VAE for Content-Based Image Retrieval. *Appl. Sci.* **2022**, *12*, 9160. [[CrossRef](#)]
41. Monowar, M.M.; Hamid, M.A.; Ohi, A.Q.; Alassafi, M.O.; Mridha, M.F. AutoRet: A Self-Supervised Spatial Recurrent Network for Content-Based Image Retrieval. *Sensors* **2022**, *22*, 2188. [[CrossRef](#)] [[PubMed](#)]
42. Kabir, M.M.; Ishraq, A.; Nur, K.; Mridha, M.F. Content-Based Image Retrieval Using AutoEmbedder. *J. Adv. Inf. Technol.* **2022**, *13*, 240–248. [[CrossRef](#)]
43. Bu, H.H.; Kim, N.C.; Kim, S.H. Content-based image retrieval using a fusion of global and local features. *ETRI J.* **2023**. [[CrossRef](#)]
44. Abdullah, S.M.; Jaber, M.M. Deep learning for content-based image retrieval in FHE algorithms. *J. Intell. Syst.* **2023**, *32*, 20220222. [[CrossRef](#)]
45. Yildirim, M.; Cinar, A. Classification of Alzheimer’s Disease MRI Images with CNN Based Hybrid Method. *Ingénierie Systèmes Inf.* **2020**, *25*, 413–418. [[CrossRef](#)]
46. Cengil, E.; Cinar, A. Multiple classification of flower images using transfer learning. In Proceedings of the 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 21–22 September 2019.
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 27–30 June 2016.
48. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [[CrossRef](#)]
49. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; Van Der Laak, J.A.; Van Ginneken, B.; Sánchez, C.I. A survey on deep learning in medical image analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [[CrossRef](#)]
50. Cinar, A.; Yildirim, M. Classification of Malaria Cell Images with Deep Learning Architectures. *Ingénierie Des Systèmes D Inf.* **2020**, *25*, 35–39. [[CrossRef](#)]
51. Jindal, I.; Nokleby, M.; Chen, X. Learning deep networks from noisy labels with dropout regularization. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016.
52. Cinar, A.; Yıldırım, M.; Eroğlu, Y. Classification of pneumonia cell images using improved ResNet50 model. *Trait. Du Signal* **2021**, *38*, 165–173. [[CrossRef](#)]
53. Magliani, F.; Sani, L.; Cagnoni, S.; Prati, A. Genetic algorithms for the optimization of diffusion parameters in content-based image retrieval. In Proceedings of the 13th International Conference on Distributed Smart Cameras, Trento, Italy, 9–11 September 2019.
54. He, T.; Wei, Y.; Liu, Z.; Qing, G.; Zhang, D. Content based image retrieval method based on SIFT feature. In Proceedings of the 2018 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Xiamen, China, 25–26 January 2018.
55. Taheri, F.; Rahbar, K.; Salimi, P. Effective features in content-based image retrieval from a combination of low-level features and deep Boltzmann machine. *Multimed. Tools Appl.* **2022**, 1–24. [[CrossRef](#)]
56. Rastegar, H.; Giveki, D. Designing a new deep convolutional neural network for content-based image retrieval with relevance feedback. *Comput. Electr. Eng.* **2023**, *106*, 108593. [[CrossRef](#)]
57. Singh, S.; Batra, S. An efficient bi-layer content based image retrieval system. *Multimed. Tools Appl.* **2020**, *79*, 17731–17759. [[CrossRef](#)]
58. Maji, S.; Bose, S. CBIR using features derived by deep learning. *ACM/IMS Trans. Data Sci. TDS* **2021**, *2*, 1–24. [[CrossRef](#)]
59. Khokhar, S.; Verma, S. Content based image retrieval with multi-feature classification by back-propagation neural network. *Int. J. Comput. Appl. Technol. Res.* **2017**, *6*, 278–284. [[CrossRef](#)]

60. Ahmed, T.; Naqvi, S.A.H.; Rehman, A.; Saba, T. Convolution, approximation and spatial information based object and color signatures for content based image retrieval. In Proceedings of the 2019 International Conference on Computer and Information Sciences (ICCIS), Aljouf, Saudi Arabia, 10–11 April 2019.
61. Sarwar, A.; Mehmood, Z.; Saba, T.; Qazi, K.A.; Adnan, A.; Jamal, H. A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine. *J. Inf. Sci.* **2019**, *45*, 117–135. [[CrossRef](#)]
62. Sikha, K.; Soman, K.P. Dynamic Mode Decomposition based salient edge/region features for content based image retrieval. *Multim. Tools Appl.* **2021**, *80*, 15937–15958.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.