ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 12 (2022) No. 2, pp. 477-496
DOI: 10.33168/JSMS.2022.0226

Image Retrieval Based on Deep Learning

Moshira S. Ghaleb, Hala M. Ebied, Howida A. Shedeed, Mohamed F. Tolba

Scientific Computing Department, Faculty of Computer and Information Sciences, Ain Shams University Cairo, Egypt

moshirasg@cis.asu.edu.eg

Abstract: The digital image has an important role in many fields such as biomedical, robotics, weather forecasting and object recognition. Due to the widespread use of social media sites, cloud services, and cellphones, large image databases are easily accessible. Searching by text is a common and simple method, however if the algorithm is running properly, searching by visual content will be much more sensitive. The goal of this research is to create a content-based image retrieval method that is more accurate in image retrieval. In this approach, intelligent systems can assist and work successfully. This study analyses three deep learning-based proposal methodologies: CNN, convolutional layers fused with LSTM, and Convolutional layers fused with GRU. The models were tested on four distinct databases of varying sizes, including Corel1K, Cifar-10, Cifar-100, and Mnist 70K. In comparison to state-of-the-art models, the three presented algorithms have significantly reduced computation time and provided very high picture retrieval levels of accuracy. For Corel1K, Cifar-10, Cifar-100, and Mnist 70k, CNN's proposed model scored 93.3, 94%, 85.5 %, and 99.9 %, respectively. the second proposed model scored 94.5%, 95%, 86.5%, and 99.9 for Corel1K, Cifar-10, Cifar-100, and Mnist 70k, respectively. Finally, for Corel1K, Cifar-10, Cifar-100, and Mnist 70k, the third proposed model reached 95.5%, 96%, 87.5 %, and 99.9%, respectively.

Keywords: CBIR, deep learning, convolution neural network, long short time memory, gate recurrent unit

1. Introduction

Digital image files are increasingly being developed and made accessible to a large number of people via the World Wide Web, rendering Image Retrieval (IR) a major research subject in computer vision. The human mind processes images considerably more quickly than words. The eye consistently provides the information it receives from seeing an image of something far faster than it does from reading a text. No one nowadays is without an online image. So, if people wish to find items for which they have no name, searching by image would be simpler. Content-Based Image Retrieval (CBIR) is a text-based image retrieval framework that uses visual qualities like color, shape, and texture as search terms.

Many strategies for improving the effectiveness of Content-Based Image Retrieval (CBIR) systems have been developed (Yu et al., 2011; El-Alami 2011; Qiu 2003; Gou et al., 2013; Lu et al., 2005). The search for items in the massive image library is one of CBIR's problems. It is simple to detect items using human eyes, but employing computers that rely on images, colors, textures, and shapes will be challenging and lead to categorization errors. The semantic gap (Pardijs) is the distinction between high-level semantics and low-level picture characteristics. Traditional methods for extracting image features for the search query and all other images in the database are used in Content-Based Image Retrieval. Histogram analysis, grid color moment (Che et al., 2013), Sobel, and Canny edge detection are just a few examples of feature selection methods. Then, using similarity measurement algorithms like Euclidean distance and Manhattan distance, compare the query image to all of the database photos.

Artificial Intelligence (AI) is the stimulation of human intelligence in computer software that aids in engaging with devices in the same way that humans engage with them. Artificial Intelligence (AI) has recently identified as the most important science of the twenty-first century, with applications in a wide range of computer vision domains. Machine learning (Krizhevsky et al., 2012; Wang 2015; Wang et al., 2016) and deep learning (Hinton et al., 2012; Shafaey et al., 2018) are two types of artificial intelligence that have attained the highest levels of accuracy in a variety of domains. AI technology could aid picture categorization in the same way that it has aided feature extraction in other disciplines such as biology, medicine (Ghaleb et al., 2021), machine learning, speech recognition, and others (Ebied 2012; Haque et al., 2018; Ayon et al., 2020). In biological pictures, AI technology has achieved high accuracies, particularly in the identification of breast, brain, and skin cancers, as well as many viral diseases (Talo et al., 2019; Celik et al., 2020; Esteva et al., 2017; Yoon et al., 2020). While CBIR requires a greater level of AI to achieve high levels of searching results and performance accuracies, numerous studies have employed AI techniques to increase CBIR accuracies (Sezavar et al., 2019; Bengio et al., 2007; Zhong et al., 2015).

CNN (Deep Convolutional Neural Networks) is a deep learning technology that can be used to extract features from images and classify them (LeCun et al., 2015; Jiang 2009). CNN is based on collecting features from the data itself in many layers. In this paper The Convolutional Neural Network was used to generate three different intelligent models in this research. Each model was tested on three separate datasets in order to get the maximum accuracy while reducing model complexity. The first model employed a Convolutional Neural Network (CNN) to extract the relevant properties of the query image and all of the images in the database, then categorize the images into the appropriate categories and get the query image's associated images from the database images. The second model adopted several CNN and Long Short Time Memory (LSTM) for feature extraction and classification, and then retrieved the relevant photos for the search query. The final model utilizes a mixture of CNN and Gate Recurrent Unit for feature extraction (GRU). The three proposed models were compared to the state-of-the-art paper’s models in the paper.

The rest of this paper is planned as follows. Section 2 summarizes the related work. Section 3 explains the proposed models. Section 4 illustrates the datasets. The result is presented and discussed in section 5. Finally, Section 6 concludes the paper.

2. Related Work

The Content-Based Image Retrieval technique involves searching for images based on their content. CBIR is a straightforward method that consists of three phases: feature extraction, similarity assessment, and picture retrieval. To create the Feature Vector (FV), first extract the relevant features for the query image, such as color, texture, and edges, using image processing techniques. Second, using the same technique as in the previous step, extract the features for all of the photographs in the database. Finally, compute the similarity measurement between the query FV and all of the images FV in order to extract the photos from the database that are the most similar to the query image.

In CBIR, feature extraction is a crucial phase. The classification stage is primarily influenced by the feature vector. Colors, textures, and edges are only a few of the characteristics of images. The color histogram was widely employed in image retrieval for years (Liapis et al., 2004) but it failed to adequately characterize the image.

Gray Level Co-Occurrence Matrix (GLCM) was used by Mohanaiah et al. (2013) to extract four Texture Parameters: Entropy, Inverse Difference Moment, Angular Second Moment, and Correlation. When compared to DWT, the image compression time is significantly reduced, according to the report. For the ant identification model, Anami et al. (2010) used a mix of color and texture features. For texture features, they employed the Sobel operator to extract the color histogram and edge direction histogram. The retrieved features were then trained using the (RBENN) network and an SVM. In comparison to SOFM, Ghaleb et al. [30] used a mixture of SOFM and

MLP to increase recognition accuracy. The paper's average recognition accuracy was around 99 %.

Deep learning has many techniques that achieved good accuracies in image classification such as CNN, LSTM, and GRU would affect the efficiency of CBIR performance. Image classification using CNN has a highly accuracies in biomedical pictures, image categorization using CNN has good accuracy. Ghaleb et al (2021) presented a CNN model for detecting Covid-19 in x-ray pictures. Two studies were used to divide the x-ray images into three categories: covid19, phenomena, and regular x-ray chests. The model classified covied19, phenomena, and normal chests with an average accuracy of 96.8%. Islam et al. (2020) is using a Convolutional Neural Network (CNN) and a Long Short Term Memory (LSTM) to automatically diagnose COVID-19 from X-ray pictures. The experiment used a total of 4575 X-ray images, 1525 of which were from the COVID-19 dataset. The AUC is 99.9%, the accuracy is 99.4%, the sensitivity is 99.3 %, the specificity is 99.2 %, and the F1-score is 98.9 %.

Ghaleb et al (2021) introduced a CNN model that measures the retrieval accuracy of 10 object images and 10 digit images with 92.9 and 99.8% average accuracy, respectively. Tan et al. (2020) utilised three distinct Convolutional Neural Network (CNN) models, namely pre-trained AlexNet, fine-tuned AlexNet, and D-Leaf, to extract features. Five machine learning algorithms were used to classify these features: Support Vector Machine (SVM), Artificial Neural Network (ANN), k-Nearest-Neighbour (k-NN), Nave-Bayes (NB), and CNN. In comparison to AlexNet (93.26%) and fine-tuned AlexNet (95.54%) models, the D-Leaf model obtained a testing accuracy of 94.88%.

You et al. blended CNN and GRU. They approach the challenge of licence plate recognition as a series modelling problem. The CNN was utilised to extract image features, and the GRU neural network was used as the sequence learning device to effectively represent the sequence's internal interactions. The accuracy of the test result is 99 %. Ghaleb et al (2022) proposed a method for classifying weather images using a CNN model mixed with DT and SVM. The model is used to determine how effective CNN is at picture classification. For the CNN model, CNN+DT model, and CNN+SVM model, the average accuracy was 92%, 93 %, and 94 %, respectively.

3. Content based Image Retrieval Proposed Models

Researchers presented Content based image retrieval approaches based on deep learning. The intelligent CBIR models consist of two phases; the training phase and the retrieve phase. The training phase used deep learning model such as CNN to extract the important features such as color, edges and textures to create the feature vector then classify the images into categories according to the extracted features the second phase is the retrieval phase. The retrieval phase retrieves the relevant images to the query image from the database and evaluates the model performance using evaluation metrics. This section presents three different deep learning models for CBIR's and the retrieve matrices.

The use of Deep Learning in CBIR will alter the CBIR strategy. The three phases of the intelligent CBIR technique are feature extraction, image classification, and image classification, which classify the images into classifications. The test phase forecasts the category for the query image and then evaluates whether the forecast is true or not, calculates the model's average accuracy. The structure of the CBIR intelligent approach is shown in Fig. 1. Training and testing are the two main processes in deep learning. The training phase is used to teach the model how to extract significant features and build the FV, after which it is used to categorize each image into one of several categories based on its FV. Calculating the classification error is critical for determining whether or not the categorization is correct. As a result, the model can continue the training until the least loss accuracy is achieved. This is something the system does for many epochs. Finally, the test phase is used to evaluate the model's performance by using some previously unseen photographs as a query image and collecting visual content from the dataset based on the categorization class. The picture retrieval measures are then used to determine how many relevant photos were retrieved. Deep learning techniques such as CNN, LSTM, and GRU, which have achieved high accuracies in image classification, might have an impact on CBIR performance. The goal of this research is to find the most intelligent deep learning strategy for solving CBIR difficulties.

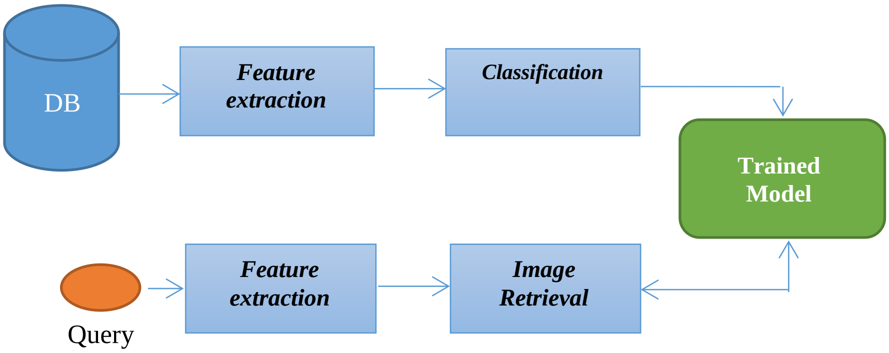


Fig. 1: Block diagram Block diagram of intelligent CBIR approach

3.1. Deep learning proposed models

3.1.1. Convolution neural network (CNN)

Yann LeCun[37] is the one to introduce convolutional neural networks in the 1980s. CNN stands for Convolutional Neural Network, which is a type of neural network designed to handle data in the form of a 2D matrix, such as pictures. CNNs are commonly employed in the detection and categorization of images.

For image recognition and classification, CNN is a powerful tool. As a result, in addition to powering vision in robots and self-driving cars, it can be used to recognize faces, objects, and traffic signs. In convolutional neural networks, the major building elements are convolutional layers. Convolution's primary function is to extract features from an input image by down sampling it into a features map utilizing information from nearby pixels. To accomplish its goal, it employs filters of a smaller size than the input size, which it combines to the picture matrix to generate the feature map. By learning visual attributes, convolution saves the spatial link between pixels.

The Convolutional Neural Network is employed in the first proposed model. The feature extraction is created by the first six layers of the CNN layers, which have 64 kernels with (11x11) kernel size, 128 kernels with (3x3) kernel size, 256 kernels with (3x3) kernel size, 256 kernels with (3x3) kernel size, 512 kernels with (3x3) kernel size, 512 kernels with (3x3) kernel size, After the second and fourth CL layers, there are two max pooling layers.

Following the six CL layers, there are three FC layers, the first of which has 1024 nodes and the second of which has 512 nodes. The Relu activation function is used by both FC levels. The third FC layer is the output layer, which has the same number of nodes as the dataset categories. The Softmax activation function is utilized in the last FC layer. The CNN's setup is show+n in Table 1.

Table 1: Setup of CNN's Model

Layer	*Filter size*	No. Of Kernel	No. of Nodes	Size of Stride	Activation Function
CL1	11x11	64	*	4x4	Relu
CL2	3x3	128	*	1x1	Relu
Max pooling	2x2	*	*	*	*
CL3	3x3	256	*	1x1	Relu
CL4	3x3	256	*	1x1	Relu
Max pooling	2x2	*	*	*	*
CL5	3x3	512	*	1x1	Relu
CL6	3x3	512	*	1x1	Relu
FC1	*	*	1024	*	Relu
FC2	*	*	512	*	Relu
FC3 (output)	*	*	10	*	softmax

3.1.2. Long short time memory (LSTM)

Sepp Hochreiter and Jürgen Schmidhuber (1997) developed Long Short Term Memory networks (LSTMs). The LSTM neural network is a type of RNN. Because there can be lags of undetermined duration between critical occurrences in a time series, it is utilized for classifying, analyzing, and making predictions based on time series data. The LSTM is made up of three gates: input, output, and forget. The three gates pass on the information into and out of the cell, and the gate remembers values across arbitrary time intervals. It can process not only single data points (like photos), but also complete data sequences (such as speech or video). Tanh and sigmoid functions are utilized as activation functions in LSTM networks.

The second proposed model is a CNN-LSTM hybrid model. In image classification, CNN produced great accuracies. It has a completely connected layer that provides learn features from all combinations of the preceding layer's features. LSTM, on the other hand, is a sort of recurrent neural network that has achieved excellent accuracy in picture classification by learning the order dependence between objects in a sequence. To achieve high picture classification accuracies, this paper combines the two models.

The proposed CNN+LSTM model uses the CNN for feature extraction and the LSTM for classification. The CNN+LSTM model architecture is shown in Table 2. The convolution layers, pooling layer, and the first two fully connected layers make up CNN's feature extraction section. After that, we add two LSTM layers with 64 and 100 nodes that use sigmoid activation function followed by fully connected layer with SoftMax function.

3.1.3. Gated recurrent unit (GRU)

Kyunghyun Cho et al. (2014) invented GRU in 2014. The GRU is a type of RNN. It has a similar appearance to LSTM; however it contains fewer parameters and gates. Only two gates make up GRU: reset and update gates. GRU contains fewer parameters than LSTM, but it is faster to perform and learn. The update gate is in charge of calculating how much past data (prior time steps) needs to be passed along to the next state. The reset gate, on the other hand, is employed by the model to determine how much past data should be ignored.

The third proposed model is a CNN and GRU mixed model. GRU is a type of RNN that has the unique capacity to recall values over time. The paper proposes a CNN feature extraction model and a GRU classification model for the feature extraction phase. The CNN+GRU model architecture is shown in Table 3. The convolution layers, pooling layers, and the first two fully connected layers make up CNN's feature extraction section. We then add two GRU layers and a fully connected layer with the SoftMax algorithm.

Table 2: Setup of CNN+LSTM‘s model

Layer	*Filter size*	No. Of Kernel	No. Of Nodes	Size of Stride	Activation Function
CL1	11x11	64	*	4x4	Relu
Max pooling	2x2	*	*	*	*
CL2	3x3	128	*	1x1	Relu
CL3	3x3	265	*	1x1	Relu
CL4	3x3	265	*	1x1	Relu
Max pooling	2x2	*	*	*	*
CL5	3x3	512	*	1x1	Relu
CL6	3x3	512	*	1x1	Relu

FC1	*	*	1024	*	Relu
FC2	*	*	512	*	Relu
LSTM1	*	*	100	*	Sigmoid
LSTM2	*	*	64	*	Sigmoid
FC3 (output)	*	*	10	*	SoftMax

Table 3: Setup for CNN+GRU‘s model

Layer	*Filter size*	No. Of Kernel	No. Of Nodes	Size of Stride	Activation Function
CL1	11x11	64	*	4x4	Relu
Max pooling	2x2	*	*	*	*
CL2	3x3	128	*	1x1	Relu
CL3	3x3	256	*	1x1	Relu
CL4	3x3	256	*	1x1	Relu
Max pooling	2x2	*	*	*	*
CL5	3x3	512	*	1x1	Relu
CL6	3x3	512	*	1x1	Relu
FC1	*	*	1024	*	Relu
FC2	*	*	512	*	Relu
GRU 1	*	*	100	*	Sigmoid
GRU 2	*	*	64	*	Sigmoid
FC3 (output)	*	*	10	*	SoftMax

3.2. Evaluation metrics

There are many well-known parameters for assessing the results of CBIR in order to make a valid comparison between the various competitive techniques. Precision and recall have long been the most prevalent evaluation parameters for CBIR. They assign an actual number between 0 and 1, with the greater the number, the better. The following equations give precision and recall:

$$\text{Precision} = \frac{N_r}{N_t} \tag{1}$$

$$\text{Recall} = \frac{N_r}{N_k} \tag{2}$$

Where, N_r is the number of true positive relevant images retrieved, N_t demonstrates total number of images retrieved (number of true positives and number of false positives) and N_k is total number of relevant images in database (number of true positives and number of false negatives). The precision metric is usually calculated to measure the ability of the system to retrieve only the images that are relevant to the query images when the number of retrieved images is k. The Recall is the ratio of the number of relevant examples retrieved to the total number of instances in the dataset for that class.

Combinations of precision and Recall metrics are used to calculate the average precision AP, P (1), and the mean average precision (mAP). They are defining as:

1. P (1): precision at 100% recall.

It calculates precisions after retrieving all the relevant images in the database for each query (i.e. retrieve all the relevant images of the suggested class).

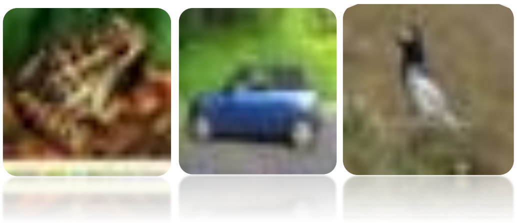
2. AP: averages the precision values; where a relevant image is retrieved for each class.
3. mAP: mean average precision; measured by computing the mean of the average precision for all the classes in the database.

4. Datasets

We utilize four separate datasets throughout the studies, each with its own collection of categories and image types. The proposed models divide the datasets into training and test images, with 80 % of the training photos and % of the test images. The first dataset is Corel 1K [40], which contains 1000 photos divided into ten categories. Figure 3 shows more Corel1k sample images.

Fig. 2: Sample of Corel1K dataset

CIFAR-10 [41] is the second dataset. It's a public dataset with 60000 photos divided into ten categories. The image size is 32x32 pixels. There are 6000 photos in each class. Cifar10 is divided into two parts: a training phase with 50000 photos and a testing phase with 10,000 images.



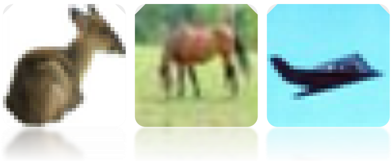


Fig 3. Sample of Cifar-101K dataset

Cifar-100[41] is the third. It's similar to the CIFAR-10, only it features 100 classes, each with 600 photos. Each class has 500 training photos and 100 assessment images. The second dataset is the MNIST [26] database, which is the fourth dataset.

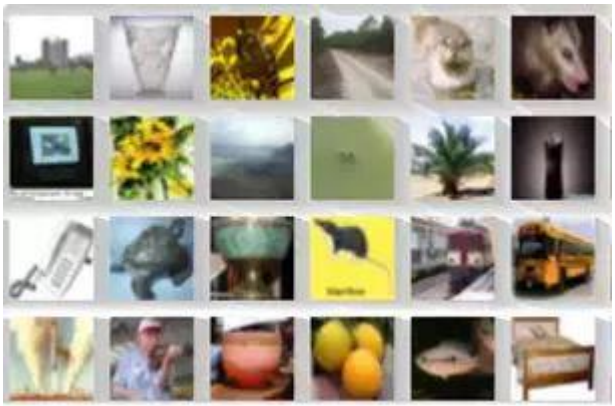


Fig. 4: Sample of Cifar-100 dataset

The fourth dataset is Mnist 70k [42] .It is a collection of 70000 handwritten character recognition images divided into ten categories. The digits 0 to 9 are divided into categories. With a 28 X28 image size, MNIST images are transformed to grayscale. The Mnist dataset is divided into two sets of 10,000 and 60,000 images, which are used for testing and training, correspondingly. The Mnist database, which is a subclass of a large database, is provided by the NIST database. All digit photos have been shrunk and bundled in a uniform size.

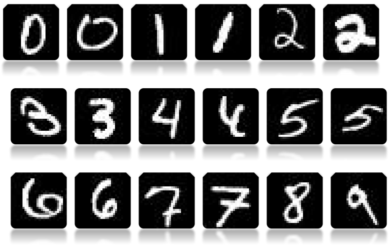


Fig. 5: Sample of Mnist dataset

5. Experimental Results

Five experiments were applied to determine the performance of the proposed models. The first experiment applied CNN models on corel1k which is small dataset to determine the suitable trainable parameter numbers for CNN. The results are compared with the Alex pre-trained network [43] using Corel 1K (10 classes). The proposed model used only 5.5M weight parameters during the training phase while Alex used 28M. The training time for the proposed model is 10s, which is less than Alex training time which is 12s. Table.4 shows the performance of CNN proposed model, Alex network, and fine-tuned Alex [44].

Table 4: Performance of CNN proposed model and state-of-the-art models

	Alex[43]	Fine-tuned Alex[44]	Proposed CNN
Trainable parameter	28 M	8M	5.5M
Batch-size	64	64	64
Epoch	50	50	50
Time S/E	14s	12s	8s
Training Accuracy	95%	96%	98%

The proposed CNN model achieved 93.3% mAP when applied on Corel1K. Fig. 6 shows mAP accuracies for the ten classes of corel1K. The result compared with the state-of-the-art models; Baig et al [52], Sarwar et al [53], and Yousuf et al [54]. We observed from this experiment that the mAP for the proposed CNN models archived the highest value competed to the state-of-the-art models. Table 5 shows the performance comparison of the CNN proposed model with the state-of-the-art models using the Mean Average Precisions (mAP) over the Corel1K dataset. Figure 7 shows the top retrieved 5 images for Corel1K dataset.

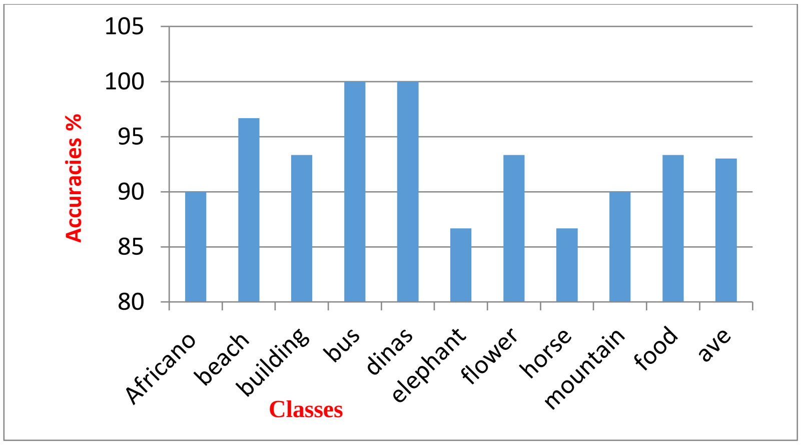


Fig. 6: The mAP used Corel1K dataset

Table5. m|AP comparison for Corel1K dataset.

Method	*mAP %*
CoHOG, SURF [52]	89
SVM[53]	89.59
K-mean, SVM [54]	91.2
CNN Proposed Model	93.3

Fig. 6: The mAP used Corel1K dataset

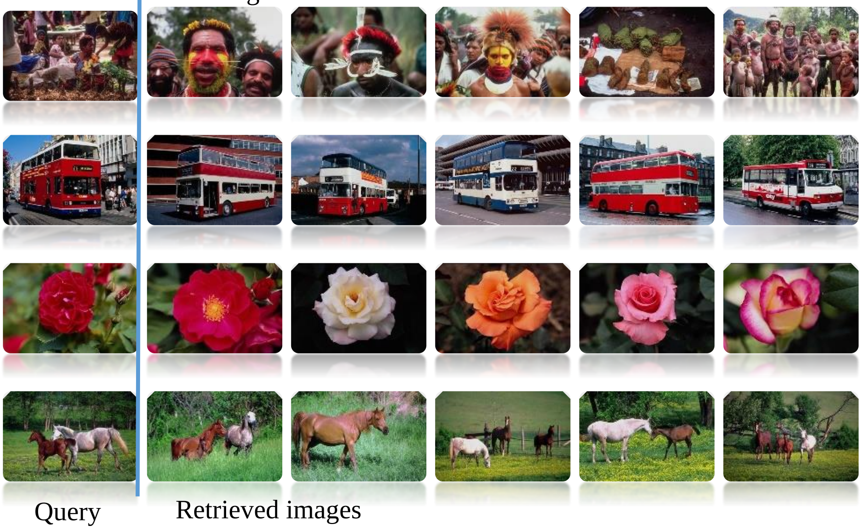


Fig. 7. Top 5 image retrieval results for the Corel1K dataset

The second experiment applied CNN model on Cifar-10 dataset. The model trained 70 epochs, used AdaMax optimization function, and used Categorial_crossentropy loss function. The model achieved 94% mAP accuracy.

Figure 8 shows the performance for CNN proposed model using the Mean Average Precisions (mAP) over the Cifar10 dataset.

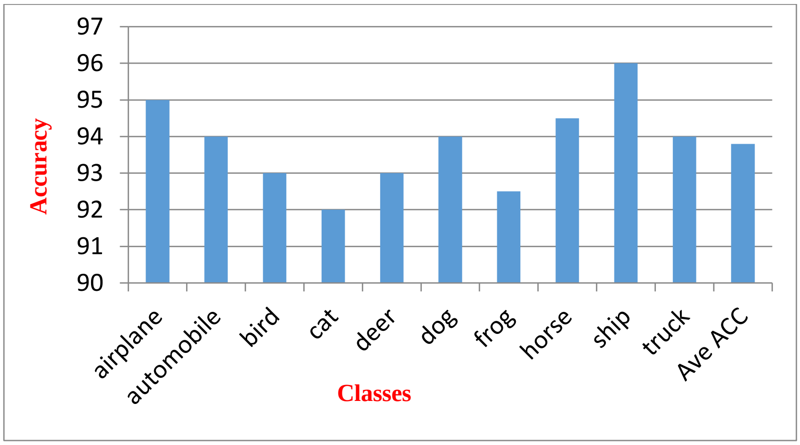


Fig. 8: The mAP used Cifar-10 dataset.

The result compared with the state-of-the-art models; Krizhevsky et al [43], Lin et al. [45], and Yang et al [46]. Ghaleb et al [33]. We observed from this experiment that the mAP for the proposed CNN models archived the highest value competed to the state-of-the-art models. Table 5 shows the performance comparison of the CNN proposed model with the state-of-the-art models using the Mean Average Precisions (mAP) over the Cifar10 dataset. Figure 9 shows the top retrieved 5 images for cifar-10 dataset

Table 5: mAP comparison for Cifar-10 dataset

Method	mAP %
Alex [43]	89
NIN + Dropout[45]	89.59
NIN + Dropout + Augmentation[45]	91.2
Hierarchical Deep ,48 nodes later layer [46]	89.6
CNN model [33]	92.9
CNN Proposed Model	94

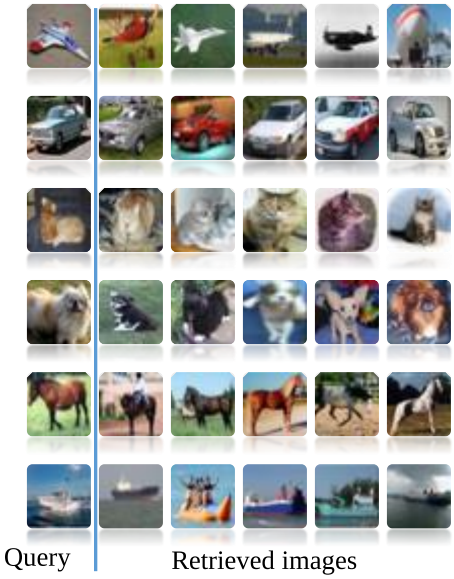


Fig. 9: Top 5 image retrieval results for the Cifar-10 dataset

The third experiment was applied the proposed CNN model on Cifar-100 dataset. The model trained 100 epochs, used AdaMax optimization function, and used Categorial_crossentropy loss function. The model achieved 85.5% mAP accuracy. The result compared with the state-of-the-art models; Zhang et al [47], Han et al. [48], and Braz et al. [49]. We observed from this experiment that the mAP for the proposed

CNN models archived the highest value competed to the state-of-the-art models. Table 6 shows the performance comparison of the CNN proposed model compared with the state-of-the-art models using the Mean Average Precisions (mAP) over the Cifar100 dataset.

Table 6: mAP COMPARISON for Cifar100 dataset

Method	*mAP %*
ResNet-110w [47]	76.95
PyramidNet [48]	81.44
Semantic Embeddings (LCORR+CLS) [ours][49]	80.94
CNN Proposed Model	85.5

The fourth experiment was applied CNN model on Mnist dataset. The classification accuracy rate is 99.9% in 30 epochs. Figure 10 shows the performance of the CNN proposed model using the Mean Average Precisions (mAP) over the Mnist dataset.The results are compared by the state-of-the-art models; Zeiler et al.[50], Chen et al. [45], Yang et al [46], Mirza et al[51], and Ghaleb et al [33]. We observed from this experiment that the mAP for the proposed CNN models archived the highest value competed to the state-of-the-art models. Table 7 shows the classification error of the CNN proposed model compared with the state-of-the-art models using the Mean Average Precisions (mAP) over the Mnist dataset. Figure 11 shows the top retrieved 5 images for Mnist dataset.

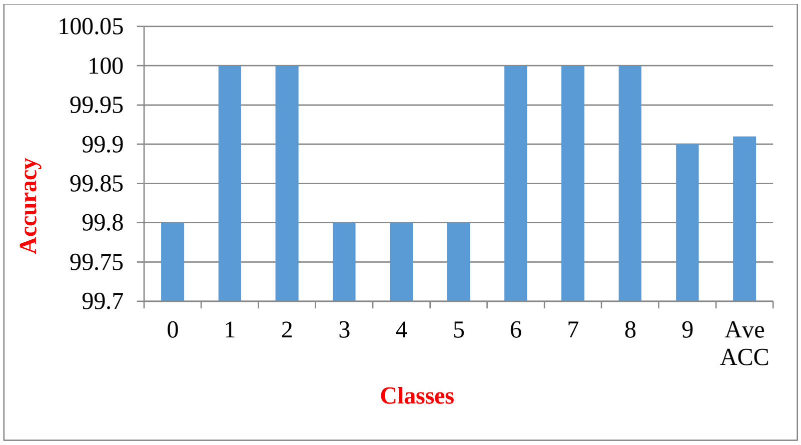


Fig. 10: The mAP used Mnist dataset table

Table 7: Classification error comparison over Mnist dataset

Method	*Test Error*
CNN + NN [50]	0.53
Stochastic Pooling [50]	0.47
NN+Dropout[45]	0.47

Conv. maxout + Dropout [51]	0.45
Hierarchical Deep [46],48 node later layer	0.47
CNN model [33]	0.2
CNN Proposed Model	0.1

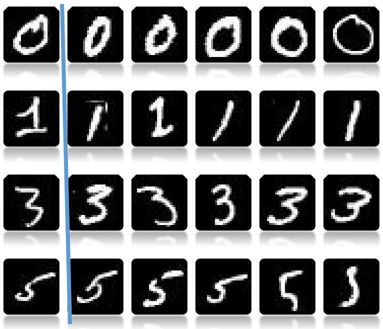


Fig. 11: Top 5 image retrieval results for the Mnist dataset

The last experiment was the proposed CNN+LSTM and CNN+GRU models. CNN+LSTM achieved 94.5%, 95%, 86.5%, and 99.9% on Corel1K, Cifar-10 and Cifar-100, and Mnist datasets. CNN+GRU achieved 95.5%, 96%, 87.5%, and 99.9% on Corel1K, Cifar-10 and Cifar-100, and Mnist datasets. The combination of CNN with LSTM and GRU achieved better accuracies compared by the state-of-the arts; Chen et al. [45], Yang et al [46], Ghaleb et al [33], and Han et al. [48]. We observed that the fusion of CNN with LSTM and GRU improve the classification accuracies and give better results in image retrieval algorithms. Table 8 shows the performance comparison of the CNN, CNN+LSTM, and CNN+GRU proposed models compared with the state-of-the-art models using the Mean Average Precisions (mAP) over the four different scales datasets ; Corel1k, Cifar-10, Cifar-100, and Mnist70k.

Table 8: mAP of the proposed models and state-of-art models

Method	mAP % for Dataset			
	Corel1K	Cifar-10	Cifar-100	Mnist 70k
NIN + Dropout + Augmentation[45]	--	91.2	--	--
Hierarchical Deep ,48 nodes later layer [46]	--	89.6	--	53
PyramidNet [48]	--	--	81.44	--
CNN model [33]	--	92.5	85.5	99.8
SVM[53]	89.59	--	--	--
K-mean, SVM [54]	91.2	--	--	--
CNN Propose model	93.3	94	85.5	99.9
CNN+LSTM Propose model	94.5	95	86.5	99.9
CNN+GRU Propose model	95.5	96	87.5	99.9

6. Conclusion

Content based image retrieval is very important today because of the huge rapid in multimedia technology.

Researches move towered create intelligent retrieval models. Deep learning has achieved great performance in computer vision. Convolution neural network has achieved high accuracy in images feature extraction and classification. The paper has used deep learning algorithms to improve the CBIR performance. This paper presents CNN, CNN fused with LSTM, and CNN fused with GRU as three different models and compares the results with the state-of-the-art models. Five experiments are presented by the paper.

The first experiment applied CNN model on Corel1K dataset to measure the model performance. The models have achieved high performance compared with the state-o- the-art models. The second experiment carried out 94% mAP when applied CNN model on Cifar-10 dataset. The results compared to the state-of-the arts and achieved the highest accuracies. The third experiment carried out 85.5% mAP when applied CNN model on Cifar-100 dataset. The results achieved the highest accuracies in comparison with the state-of-the-art models. The fourth experiment carried out 99.9% mAP when applied CNN model on Mnist dataset. The results compared by the stat-of-the-art models and achieved the highest accuracies.

The last experiment applied a fused model of Convolution and LSTM layers on Corel1k, Cifar-10, Cifar-100, and Mnist datasets, respectively. Additionally, the experiment applied a confusion model of convolution and GRU on Corel1k, Cifar-10, Cifar-100, and Mnist datasets, respectively. The experiment achieved the highest accuracies in comparison with the stat-of-the-art.

References

Anami, B. S., Suvarna, S. N., & Govardhan, A. (2010). A combined color, texture and edge features based approach for identification and classification of Indian medicinal plants. *International Journal of Computer Applications,* 45-51.

Ayon, S. I., Islam, M. M., & Hossain, M. R. (2020). Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. *IETE Journal of Research,* 1-20.

Baig, F., Mehmood, Z., Rashid, M., Javid, M. A., Rehman, A., Saba, T., & Adnan, A. (2020). Boosting the performance of the BoVW model using SURF-CoHOG-based sparse features with relevance feedback for CBIR. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering,* 44(1), 99-118. DOI:https://doi.org/10.1007/s40998-019-00237-z.

Barz, B. & Denzler, J. (2019). Hierarchy-based image embeddings for semantic image retrieval. *2019 IEEE Winter Conference on Applications of Computer Vision (WACV),* 638-647. DOI:10.1109/WACV.2019.00073.

Bengio, Y., Lamblin, P., Popovici, D., & Larochelle, H. (2007). Greedy layer-wise training of deep networks. *In Advances in Neural Information Processing Systems,* 153-160.

Celik, Y., Talo, M., Yildirim, O., Karabatak, M., & Acharya, U. R. (2020). Automated invasive ductal carcinoma detection based using deep transfer learning with whole-slide images. *Pattern Recognition Letters*, 133, 232–239.

Che, H. N. A., Jamil, N., Nordin, S., & Awang, K. (2013). Plant species identification by using Scale Invariant Feature Transform (SIFT) and Grid Based Colour Moment (GBCM). *IEEE Conference on Open Systems (ICOS),* Kuching, 226-230.

Cho, K., van Merrienboer, B., Gulcehre, C., Bahdanau, D., Bougares, F. Schwenk, H., & Bengio, Y. (2014) Coral dataset, last referred on June 2009, Available at http://wang.ist.psu.edu/docs/related/.

Ebied, H. M. (2012). Feature extraction using PCA and Kernel-PCA for face recognition. *8th International Conference on Informatics and Systems (INFOS),* Cairo, MM-72-MM-77.

El Alami, M. E. (2011). A novel image retrieval model based on the most relevant features. Knowledge-Based Systems. 24(1), 23-32.

Esteva, A., Kuprel, B., Novoa, R. A. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature,* 542, 115-118.

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2021). COVID-19 X-rays model detection using convolution neural network. *Artificial Intelligence and Computer Vision,* AICV, 1377, Springer.

Ghaleb, M. S., Ebied, Shedeed, H. A., & Tolba, M. F. (2019). Image retrieval based on self-organizing feature map and multilayer perceptron neural networks classifier. *Ninth International Conference on Intelligent Computing and Information Systems (ICICIS),* Cairo, Egypt, 189-193.

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2021). COVID-19 X-rays model detection using convolution neural network. *International Conference on Artificial Intelligence and Computer Vision (AICV2021), Springer International Publishing,* 3-11, Morocco.

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., & Tolba, M. F. (2021). Content based image retrieval based on convolutional neural network. *10th International Conference on Intelligent Computing and Information science (ICICS),* 149-153, Cairo, Egypt.

Ghaleb, M. S., Ebied, H. M., Shedeed, H. A., Tolba, M. F. (2022). Weather classification using fusion of deep convolutional neural networks and traditional classification methods. *IJICIS Journal*, Cairo, Egypt.

Goodfellow, I. J., Warde-Farley, D., Mirza, M., Courville, A., & Bengio, Y. (2013). Maxout networks. arXiv preprint arXiv: 1302.4389.

Guo, J. -M., Prasetyo, H., & Su H. -S. (2013). Image indexing using the color and bit pattern feature fusion. *Journal of Visual Communication and Image Representation*, 24(8), 1360-1379.

Han, D., Kim, J., & Kim, J. (2017). Deep pyramidal residual net- works. *In IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5927-5935.

Haque, M. R., Islam, M. M., Iqbal, H., Reza M. S., & Hasan, M. K. (2018). Performance evaluation of random forests and artificial neural networks for the classification of liver disorder. *International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshah, 1-5.

He, X. Zhang, S. Ren, & J. Sun, (2016). Deep residual learning for image recognition. *In IEEE Conference on Computer Vi- sion and Pattern Recognition (CVPR),* 770–778.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speechrecognition. *The shared views of four research groups. IEEE Signal Processing Magazine*, 29(6), 82-97.

Hochreiter, S., & Schmidhuber. (1997). Long short-term memory. Cambridge, MA, USA. 9, 1735-1780.

Jiang, X. (2009). Feature extraction for image recognition and computer vision. *2nd IEEE International Conference on Computer Science and Information Technology*, Beijing. 1-15.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *In Advances in neural information processing systems*, 1097-1105.

Krizhevsky. (2009). Learning multiple layers of features from tiny images. Computer Science Department, University of Toronto, Tech. Report.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*. 436-444.

LeCun, Y. & Cortes, C. (2010). MNIST handwritten digit database.

LeCun, Y., Haffner P., Bottou L., & Bengio Y. Object recognition with gradient-based learning.. In: Shape, Contour.

Liapis, S. & Tziritas, G. (2004). Color and texture image retrieval using chromaticity histograms and wavelet frames. *IEEE Trans Multimedia*, 676-686.

Lin, M., Chen, Q., & S. Yan. (2014). Network in network.

Lin, K., Yang, H., Hsiao, J., & Chen, C. (2015). Deep learning of binary hash codes for fast image retrieval. *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 27-35.

Lu, Z. -M. & Burkhardt, H. (2005). Colour image retrieval based on DCT-domain vector quantisation index histograms. *Electronics Letters*, 41(17), 956–957.

Mohanaiah, P., Sathyanarayana, P., & GuruKumar, L. (2013). Image texture feature extraction using GLCM approach. *International Journal of Scientific and Research Publications*, 3(5), 1-5.

Md. Zabirul, I., Md. Milon, I., & Amanullah, A. (2020). A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. *Informatics in Medicine Unlocked*, 20.

Pardijs, M. User-centered reduction of the semantic gap in content-based image retrieval.

Qiu, G. (2003). Color image indexing using BTC. *IEEE Transactions on Image Processing*, 2(1), 93-101. Pmid: 18237882.

Sarwar, A., Mehmood, Z., Saba, T., Qazi, K. A., Adnan, A., & Jamal, H. (2019). A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine. *Journal of Information Science*, 45(1), 117-135. DOI: https://doi.org/10.1177/ 0165551518782825.

Sezavar, A., Farsi, H., & Mohamadzadeh, S. (2019). Content-based image retrieval by combining convolutional neural networks and sparse representation. *Multimed Tools Appl*, 78, 20895–20912.

Shafaey, M. A., Salem, M. A. M., Ebied, H. M., Al-Berry, M. N., 7 Tolba, M. F. (2018). Deep learning for satellite image classification. *The International Conference on Advanced Intelligent Systems and Informatics*, 845. Springer.

Talo, M., Yildirim, O., Baloglu, U. B., Aydin, G., & Acharya, U. R. (2019). Convolutional neural networks for multiclass brain disease detection using MRI images, 78.

Tan, J. W., Chang, S. -W., Abdul-Kareem, S., Yap, H. J., & Yong, K. -T. (2020). deep learning for plant species classification using leaf vein morphometric. *In IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 17(1), 82-90.

Vanishree & Ramana, R. K. V. (2013). Implementation of pipelined Sobel edge detection algorithm on FPGA for high speed applications. *International Conference* *on Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA)*, 1-5.

Wang, Z. (2015). The applications of deep learning on traffic identification. BlackHat USA.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. ArXiv preprint arXiv: 1606.05718.

Yoon, S. H. & Lee, K. H. (2020). Chest radiographic and CT findings of the 2019 novel coronavirus disease (COVID-19): Analysis of nine patients treated in Korea. *Korean J. Radiol*, 21, 494-500.

You, F., Zhao, Yangze, & Wang, X. Combination of CNN with GRU for plate recognition. *Journal of Physics: Conference Series*, 1187, 032008, 10.1088/1742-6596/1187/3/032008.

Yousuf, M., Mehmood, Z., Habib, H. A., Mahmood, T., Saba, T., Rehman, A., & Rashid, M. (2018). A novel technique based on visual words fusion analysis of sparse features for effective content-based image retrieval. *Mathematical Problems in Engineering*, 1-13, DOI:https://doi.org/10.1155/2018/2134395.

Yu, F. -X., Luo, H., & Lu, Z. -M. (2011). Colour image retrieval using pattern co-occurrence matrices based on BTC and VQ. Electronics letters. 47(2), 100-101.

Zeiler, M. D. & Fergus, R. (2013). Stochastic pooling for regularization of deep convolutional neural networks.

Zhong, Z., Jin, L. & Xie, Z. (2015). High performance offline handwritten Chinese character recognition using googlenet and directional feature maps. *International Conference on Document Analysis and Recognition (ICDAR)*, IEEE, 846-850.