Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 02, PP. 140-152, 2025
DOI: https://doi.org/10.54216/JISIoT.140212

Enhancing Convolutional Neural Network for Image Retrieval

Zena M. Saadi[1,*], Ahmed T. Sadiq[1], Omar Z. Akif[2], El-Sayed M. El-kenawy[3,4]

[1]Computer Science Department, University of Technology, Baghdad, Iraq
[2]Department of Computer Science, College of Education for Pure Science (Ibn al-Haitham), University of Baghdad, Baghdad, Iraq
[4]Applied Science Research Center. Applied Science Private University, Amman, Jordan
[5]Jadara University Research Center, Jadara University, Jordan

Emails: cs.22.15@grad.uotechnology.edu.iq; ahmed.t.sadiq@uotechnology.edu.iq; omar.z.a@ihcoedu.uobaghdad.edu.iq; skenawy@ieee.org

Abstract

With the continuous progress of image retrieval technology, the speed of searching for the required image from a large amount of image data has become an important issue. Convolutional neural networks (CNNs) have been used in image retrieval. However, many image retrieval systems based on CNNs have poor ability to express image features. Content-based Image Retrieval (CBIR) is a method of finding desired images from image databases. However, CBIR suffers from lower accuracy in retrieving images from large-scale image databases. In this paper, the proposed system is an improvement of the convolutional neural network for greater accuracy and a machine learning tool that can be used for automatic image retrieval. It includes two phases; the first phase (offline processing) consist of two stages; stage1 for CNN model classification while stage 2 for extracts high-level features directly from CNN by a flattening layer, which will be stored into a vector. In the second phase (online processing), the retrieval depends on query by image (QBI) from the system, which relies on the online CNN model stage to extract the features of the transmitted image. Afterward, an evaluation is conducted between the extracted features and the features that were previously stored by employing the Hamming distance to return all similar images. Last, it retrieves all the images and sends them to the system. Classification for images was achieved with 97.94% deep learning results, while for retrieved images, the deep learning was 98.94%. For this paper, work done on COREL image dataset. The images in the dataset used for training are more difficult than image classification due to the need for more computational resources. In the experimental part, training images using CNN achieved high accuracy, proving that the model has high accuracy in image retrieval.

Received: March 19, 2024 Revised: June 30, 2024 Accepted: October 18, 2024

Keywords: Convolutional Neural Network (CNN); Images Classification; Retrieval images; COREL's Image Dataset; Deep Learning

1. Introduction

Image classification is one of the most important research areas of computer vision and the foundation of numerous categories of vision recognition. The enhancement of the classification network's performance usually greatly enhances its application level such as object detection, segmentation, human pose estimation, video classification, object tracking, and super resolution technology [1]. Strengthening the ability in image classification is greatly significant in advancing the cause of computer vision. Its principal activity encompasses image data pre-processing, feature extraction and feature representation, and classifier construction [2].

Image classification's main research emphasis has always been on feature extraction of images to classify images [3]. The more traditional techniques for image feature extraction pay more attention to the automatic definition of specific image features. It has low generalization performance [4]. Thus, letting a computer upload the possibility of processing images in a way close to biological vision is what researchers want. ANN is an abstract biological neural network, which is a mathematical operation model constituted by numerous neurons with a complex connection mode. It tends to mimic the neural network functionality involved in the evaluation of neural signals [5].

CBIR is the most relevant approach to useful image search because the number of digital images on digital resources has increased dramatically. Conventional image search is mostly done using the metadata of the specific image. For accessing this metadata, text-based querying methods are employed. However, to obtain better related images, all images must have correct metadata information. One of the issues affecting the accuracy of a search is the selection of the right metadata for an image. Therefore, it is much better to use what is known as the query by Image (QBI). It has been found that the CBIR system is in use for image processing searches in some research studies. CBIR system is the one that begins with a sample image and looks in a sizable database of photos for images that resemble the sample by utilizing low level features [6].

The motivation for the proposed work lies in creating a robust AI model that ensures equal accuracy and performance in image retrieval while at the same time eliminating the time it takes to complete such operations.

The following are the major contributions to this paper:

1. It must be noted that there is a trade-off between the security and the accuracy of the CBIR.

2. Providing a working strategy of image retrieval that employs the specified methods of the deep learning approach.

3. Enhancing the accuracy of the classification CNN by training it with augmented images.

4. Choose the most important features that will enhance accurate retrieval and, at the same time, reduce search time.

2. Related Works

Content-based image retrieval (CBIR) is one of the computer vision-concerned research areas, which deals with the retrieval of digital images from large databases. A basic CBIR system enables the user to input an image to search for images that have content similar to the input image. Traditional image retrieval, on the other hand, is text-based, meaning that the image's name, textual content, and indexing relationships all serve as the query function [7]. The basic concept of CBIR is the retrieval of images based on the visual features of images. In essence, it is an approximate matching technique that leverages the technological advancements of a diverse range of fields, such as databases, image processing, computer vision, and image understanding. The computer can automatically perform feature extraction and indexing, thereby avoiding the subjectivity of manual description. Retrieval mainly involves a sample image (query) being given, and then the features of the query image are extracted and matched with features in the database, finally delivering images similar to the query features to the user [8].

As the technique as seen earlier has numerous uses, deep-learning-based image retrieval has also been used on cultural heritage too. Gupta et al. [9] collected a new dataset including monument images of historical, cultural, as well as religious significance from all over India.

In Z. Cao et al. [10] reduction of data dimension and image search by convolutional neural network was applied to get higher-level features of the images. Further, to handle issues of large feature dimensions as well as strongly correlated, multiline core component analysis was applied. To speed up the computations and efficiency, the features are binary coded after the feature reduction has been performed.

Kuo et al. [11] deep convolutional neural networks were introduced in their paper for image retrieval. It uses DL for training the weight of a NN and, at the end, extracts high-level image features.

Huang et al. [12] suggested CNN as the model of an ensemble of models for image search. The feature extraction of this image classifier is accomplished by utilizing two very effective deep learning networks, namely AlexNet as well as Network in Network (NIN). It computes the working feature vectors for the weights of the images used in retrieval.

Khan and Javed [13] CBIR system was successfully developed for effective content-based search of images that would allow the system to retrieve correct images semantically. For this purpose, they propose a hybrid feature descriptor for the color and texture features.

Ma et al. [14] Constructed a dataset for application to examples of the traditional Chinese architectural heritage and proposed a deep retrieval method for the recommended task.

Medus et al. [15] apply convolutional neural networks (CNNs) to the classification of hyperspectral images, with an application in the industrial food packaging domain. Hyperspectral imaging captures and analyzes the data in the full spectrum range and discriminates the objects by their spectral responses. This can be important for maintaining product quality, sorting parts for recycling, and following food hygiene standards.

Kiranyaz et al. [16] presented a classification framework, which was named the region-based pluralistic CNN that can encode context-aware semantic representations to get promising features. Fusing several discriminant appearance factors as vector with the CNN enables to show the spatial-spectral contextual sensitivity as the need for accurate pixel classification. As for the proposed method it can be considered that when learning contextual interaction features using various region-based inputs discriminant power should be higher. Following this, the herein introduced combined representation that contains both spectral and spatial features is feed into the fully connected network in which the label of each pixel vector is determined by the Softmax layer. Based on the experimental comparison of the results of hyperspectral image datasets that are generally used in the present study, it can be found that the proposed technique is much effective relatively to other traditional deep-learning-based classifiers and many other advanced classifiers.

The authors X. Zhang et al. [17] suggested a space and class structure reconstructed sparse representation or SCSSR graph for semi-supervised HSI classification. In particular, graph Laplace has been introduced into the SR model, that is, spatial neighbors should have similar representation coefficients in the SR model, and thus, the obtained coefficients matrix is more reflective of the similarity between samples. In addition, they also include probabilistic class structure in the SR model. This corresponds to the probabilistic relationship of each sample with each class to make the structure of the graphs more unique.

3. Materials and Methods

COREL's Image Dataset

Several high-quality digital images are selected from COREL's Images database and which constitutes only a small fraction of the total Image database used to train machine learning algorithms and computer vision systems. Therefore, this dataset is becoming more and more popular for machine learning tasks. The COREL's Images dataset has a diverse range of images, covering various topics, such as landscapes, cityscapes, people, portraits, animals, wildlife, food, drink, and objects, all of which are 32×32 in size and are JPEG and TIFF formats, with a resolution of 256×384 pixels or 384×256 pixels and a maximum file size of 10 MB. So there are 10 concept groups of images composed of 100 images. For each concept group, the images are divided into 70 images for training and 10 validation with 20 for testing and evaluating the proposed model, as illustrated in Fig. 1 presents the sample images, which are taken from each semantic category of the COREL's 1000 image dataset [18].

Figure 1. Some images in subset of COREL's image database [19].

1. Image Retrieval

Image retrieval is a fundamental task in computer vision that involves searching for images in a database that match a given query image. Types of image retrieval, such as content-based image retrieval (CBIR) and Hybrid image retrieval, are retrieved based on a combination of textual and visual features [20]. Medical imaging: Image retrieval is used in medical imaging to retrieve images of medical conditions or diseases that match the query.

Image retrieval is a process that involves the following steps: [21]

- Query Image: A user provides a query image that they want to search for in the database.
- Database: A large collection of images is stored in a database.
- Feature Extraction: Features are extracted from the query image and the images in the database.
- Similarity Measurement: The similarity of the features of the query image is compared with the features of the database images.
- Ranking and Retrieval: The images in the database are ranked based on their similarity scores, and the top-ranked images are retrieved as the search results [22].

Techniques Used in Image Retrieval:

- Feature Extraction: Features are extracted from images using techniques such as convolutional neural networks (CNNs), scale-invariant feature transform (SIFT), or speeded-up robust features (SURF).
- Similarity Measurement: Similarity metrics, such as Euclidean distance or cosine similarity, are used to compare the features of images.
- Indexing and Retrieval: Indexing techniques, such as k-d trees or ball trees, are used to efficiently retrieve images from the database [23].

2. Deep Learning

In deep learning, the network seeks to identify complex features from the data with high dimensions and then use these features to build a model that maps input data to the output data (like classes) [24]. These deep learning architectures are commonly developed as neuron networks, where higher-level characteristics are estimated as nonlinear functions of lower-level characteristics. An example of a deep learning structure is the layer neural network, which is the most common in use. In its hierarchical structure, deep learning can be divided into many layers. The following section will describe these layers [25].

2.1 Convolutional Neural Network (CNN)

Convolution Neural Network is a type of neural network employed for structures with data arranged in a grid format, like images. Whose purpose is to find the attributes that are learned from the database. The convolutional layer will have is of size N x N and in this layer, dot product multiplication will be performed between the neighborhood values. Therefore, the convolutional layer only includes addition and multiplication operations in its formulation as shown in Fig. 2 [26].

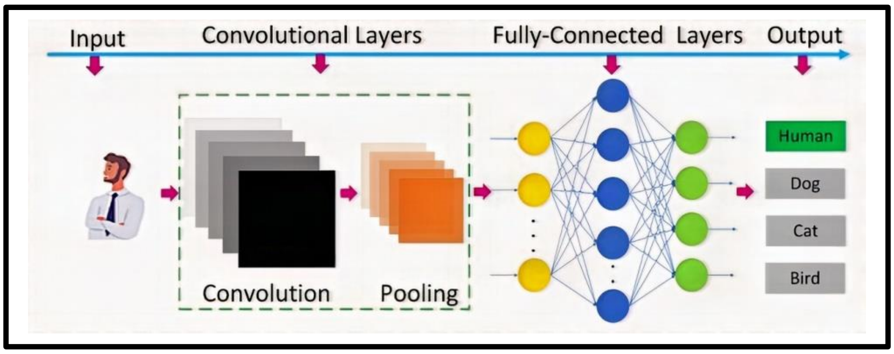


Figure 2. CNN's Internal Layers [27].

Key features of CNNs include:

1. Convolutional Layers: These layers convolve (or correlate) filters (or kernels) with the input data to find patterns such as edges and texture. This operation helps to decrease the number of dimensions with structure retention of essential characteristics. The filters are those small matrices that move over the image and carry out dot product at every point. It proceeds to emphasize particular patterns in the image.

2. Pooling Layers: Down sampling layers in a pooling layer decrease the size of the feature maps that are generated by the convolutional layers and decreases the parameter and computational complexity of the network to avoid overfitting. Some of them are max pooling, where the maximum value under the window is considered, and average pooling, where the average of the values in the window is considered.

3. Activation Functions: After convolution, a non-linear activation function is applied, such as ReLU (Rectified Linear Unit) to enable the network to learn non-linear features.

4. Fully Connected Layers: At the end of the network these layers link every neuron from the previous layer to all neurons in the current layer, commonly for classification purposes.

CNNs are well suited for image recognition, object detection, as well as numerous other computer vision tasks since they can learn features of raw data on their own [28][29] [30].

3. Proposed System

As illustrated in Fig. 3, the system uses an image and processes it through two phases to retrieve similar images from a database. The first phase that is involved is the classification model of the images and feature extraction model stages. In the second phase as implementation CBIR based on these extracted features are utilized to retrieve images from the database that are similar to the desired image as CBIR model.

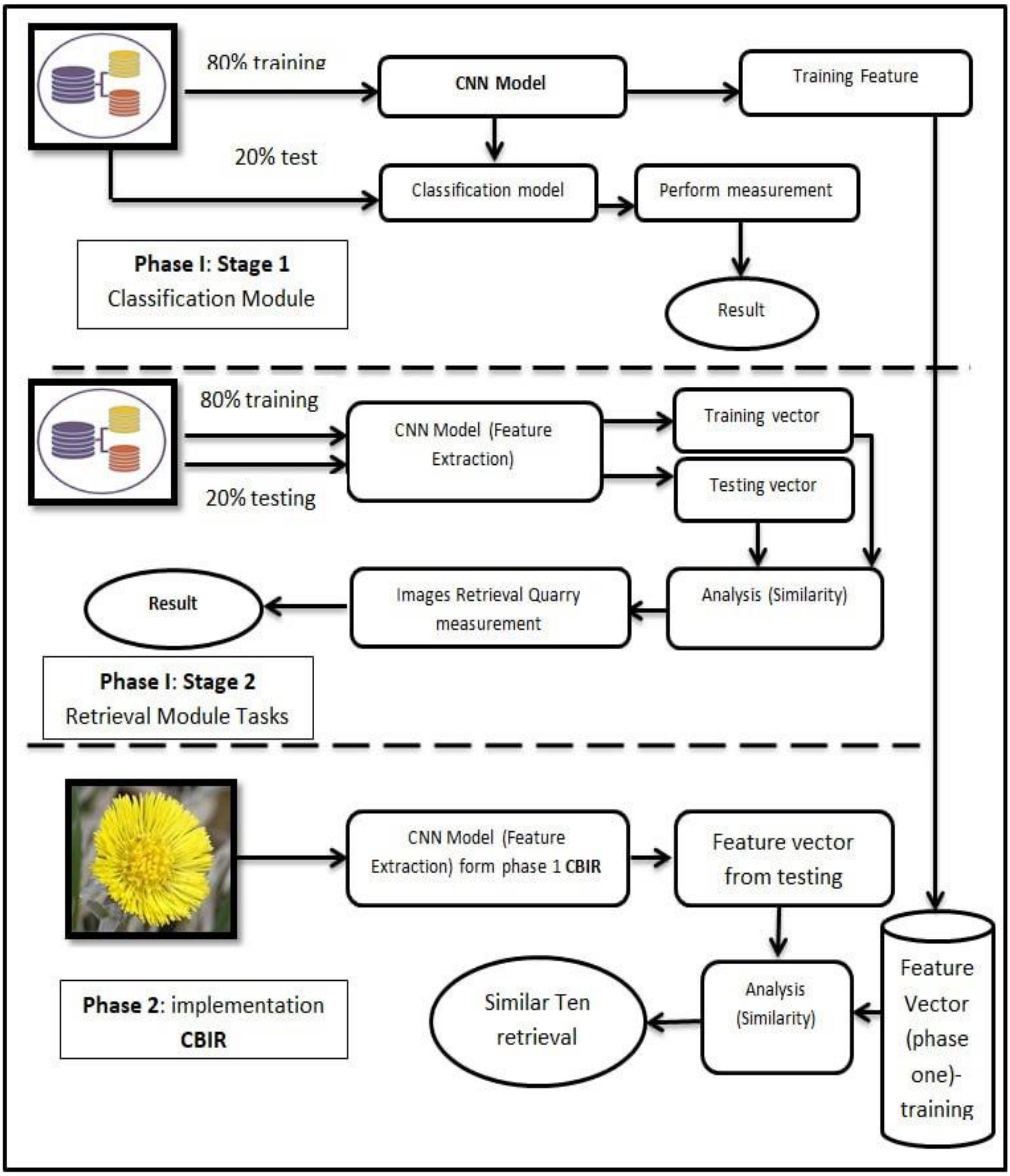


Figure 3. Proposal Retrieval Model based on customize CNN Model

Can explain the system architecture for content-based image retrieval (CBIR) that involves two main phases: classification and retrieval in details as:

3.1 Phase I: Classification

This phase is divided into two stages:

Stage 1: Classification Module Tasks

This phase consists of several procedures that are significant in the creation of the classification model. The images are classified by using a Convolutional Neural Network (CNN) model. To assess and measure the performance of the model, 20% of the images are used for testing, while the remaining 80% are employed in training the model. The outcome of this stage is a classification model.

The CNN structure of the proposed model has 10 layers, this means that each image in the training set is processed through all these layers. This network consists of a fully connected layer, five maximum pool layers, and four convolution layers. Fig. 4 shows the sequence of operations from the input to the output, which breaks it down into layers:

1. Input Layer: This layer takes the input image that usually represented as a multi-dimensional array, with dimensions corresponding to height, width, and color channels (RGB).

2. Convolutional Blocks: These blocks are the core of CNNs, responsible for extracting features from the input image. Each block typically contains: Convolutional Layers, Activation Function and Pooling Layers as shown in the following paragraphs.

a. Block 1 (Conv)

- Block 1 - Conv1: The first convolution in the block applies a kernel of size (3 x 3) with a stride of 1, using 64 filters. This extracts basic features from the input image.

- Block 1 - Conv2: The second convolution uses a kernel of (3 x 3) with a stride of 1, employing 64 filters. It builds upon the features extracted in the previous layer.

- Block 1 - Pool: A max pooling layer with a window size of (2 x 2) and a stride of 2 is applied, downsampling the feature maps.

b. Block 2 (Conv)

- Block 2 - Conv1: Similar to the first block, this convolution uses a kernel of size (3 x 3) with a stride of 1, employing 128 filters.

- Block 2 - Conv2: Another convolution with a (3 x 3) kernel, stride of 1, and 128 filters.

- Block 2 - Pool: Max pooling with a window size of (2 x 2) and stride of 2, further reducing the dimensions.

c. Block 3 (Conv)

- Block 3 - Conv1: A (3 x 3) kernel with stride 1 and 256 filters.

- Block 3 - Conv2: A (3 x 3) kernel with stride 1 and 256 filters.

- Block 3 – Conv3: A (3 x 3) kernel with stride 1 and 256 filters.

- Block 3 – Conv4: A (3 x 3) kernel with stride 1 and 256 filters.

- Block 3 - Pool: Max pooling with a window size of (2 x 2) and stride of 2.

d. Block 4 (Conv)

- Block 4 - Conv1: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 4 - Conv2: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 4 – Conv3: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 4 – Conv4: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 4 - Pool: Max pooling with a window size of (2 x 2) and stride of 2.

e. Block 5 (Conv)

- Block 5 - Conv1: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 5 - Conv2: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 5 - Conv3: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 5 - Conv4: A (3 x 3) kernel with stride 1 and 512 filters.

- Block 5 - Pool: Max pooling with a window size of (2 x 2) and stride of 2.

3. Global Max Pooling: This layer aggregates the spatial information from the last convolutional block, reducing the feature maps to a single value per feature. It takes the maximum value within each feature map, summarizing the key information learned by the network in the proposal structure using the five pooling and one max pooling function.

4. Flattening: The output of the global max pooling is flattened, converting it into a one-dimensional vector. This is necessary for feeding the data into the fully connected layers.

5. Fully Connected Layers: These layers perform the classification task by taking the flattened feature vector and processing it to predict the class labels.

- Dense Layers: These layers apply a linear transformation to the input, followed by an activation function in proposal structure for CNN us using two dense function layers.

- Dropout Layer: This layer randomly disables a fraction of neurons during training, preventing overfitting.

6. Output Layer: The final layer of the network, which outputs the predicted class probabilities.

- Activation Function: Typically, a softmax function is used, which ensures that the output probabilities for all classes sum to 1.

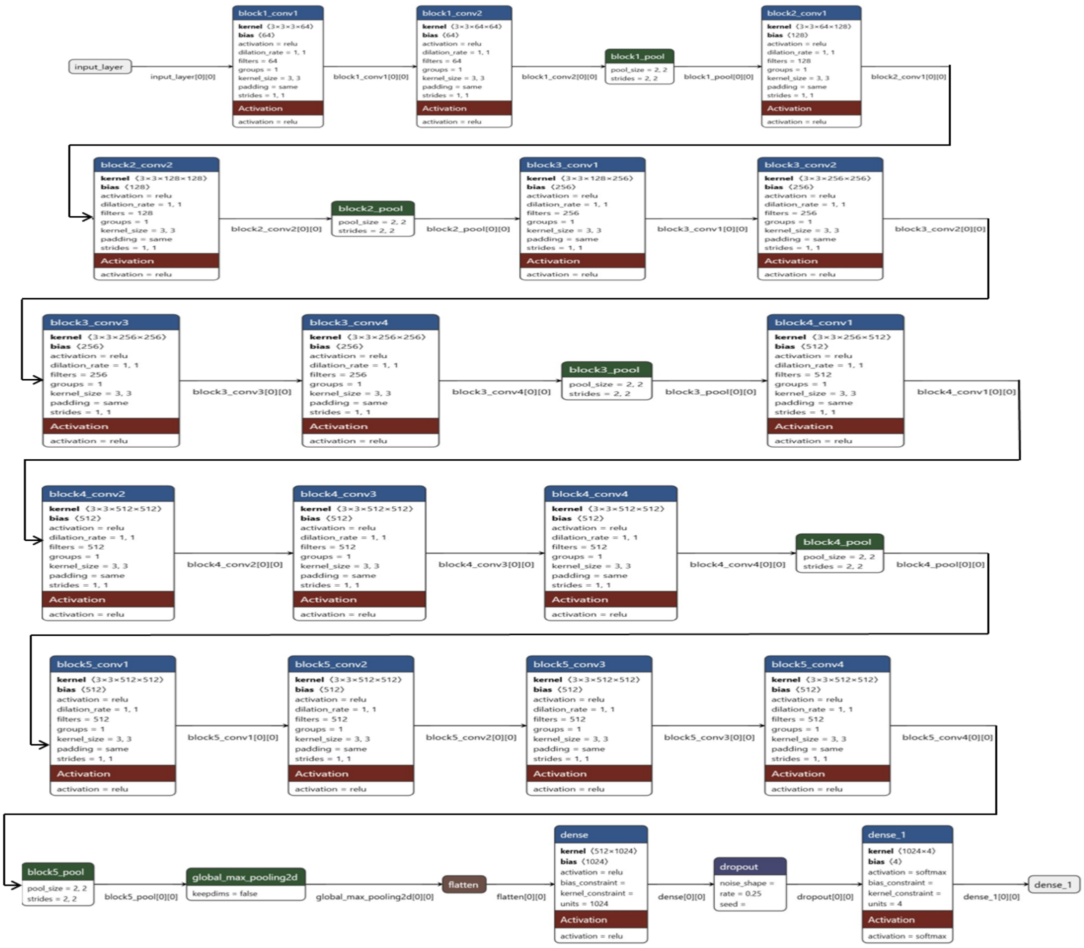


Figure 4. Proposed CNN architecture in phase I drawn by Netron software.

The proposed system **Parameters of the proposal CNN**

is a CNN structure, which is used to predict and extract high-level features. This model comprises a range of parameters that improve the efficiency of the proposed system, as depicted in Table 1.

Table 1: CNN algorithmic parameters

Parameter Name	Parameter value	Description
Kernal	3x3	This filter is used to extract features from images.
Padding	Same	Determines how the input is padded before convolution (often "same" to maintain the same spatial dimensions).
Dropout	%25	Randomly drops out a portion of input neurons to zero during training to prevent overfitting.
Pooling(Maxpooling)	2x2	Select the maximum value from a defined region of the feature map is taken. This retains important features while reducing dimensionality.
Epochs	10	The COREL's Images dataset is processed through the neural network 10 times.
batch size	32	The number of input data samples that the model takes as input in one single update.
weight decay	0.0001	Used to prevent overfitting by adding a penalty term to the loss function for large weight values.
Loss function	Categorical crossentropy	To quantify the error between the predicted output and the desired output, so that the model can learn to minimize this error during training.

Stage 2: Retrieval Module Tasks

To extract the features from the images, the system employs CNN model specialized for feature extraction by using the output of model flattening, which turns all the received two-dimensional matrices into a line of vectors. It is through these features that training and testing vectors can be created. For each of the training images, it generates 1024 features and provides the feature vector for all the training images in the form of a matrix. By using some similarity analysis method, the system then matches up these extracted features with features stored in a database. This stage aims to identify similar images based on extracted features.

3.2 Phase II: Implementation of CBIR

This phase implements the CBIR system using the feature vectors obtained in the preceding phase; subsequently, the system retrieves the ten closest images in the database based on the feature vectors extracted from the images based on phase 1 in training feature vectors. This retrieval process is constructed on a similarity comparison of the query image with images in the database. Where the same proposal model (CNN) was used without the flatting stage and this can be clarified in the following steps:

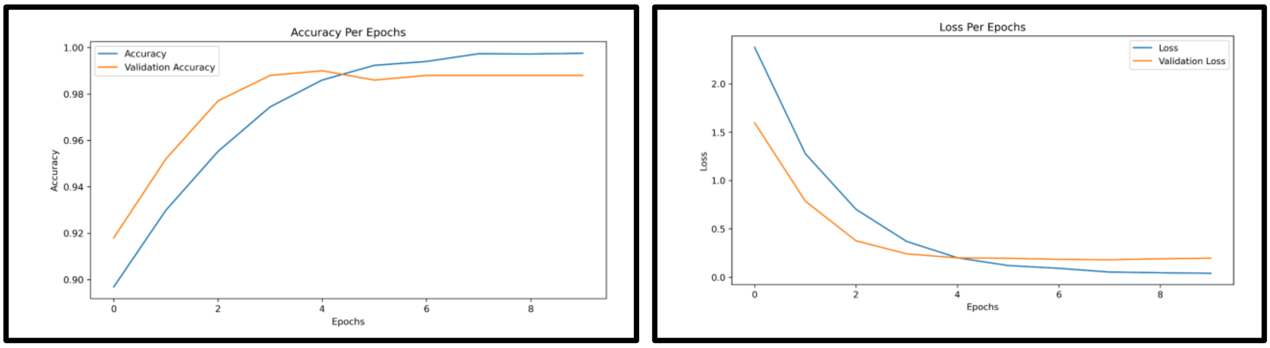
1. Extract Features: Pass each preprocessed database image through the same proposal CNN model to extract features.
2. Get the Feature Vectors: The output of the CNN is a feature vector for each database image.
3. Calculate Similarity: Calculate the similarity between the query image feature vector and each database image feature vector using a similarity metric (EUclidean distance).
4. Get the Similarity Scores: The output of the similarity calculation is a similarity score for each database image.
5. Rank the Database Images: Rank the database images based on their similarity scores.
6. Retrieve the similarity Images: Retrieve the Ten images with the highest similarity scores.

4. System Analysis

Many of the assumptions of machine learning are based on visualizations. Both professionals and scholars often use visualization to control the learned parameters and the output results to enhance their models. Besides graph visualization, there is a module for displaying the distribution of tensors images, and sounds in the TensorBoard component of the dashboard. In this paper, analysis of a two-phase CNN proposal and retrieval model.

4.1 Phase I: Stage 1 (CNN MODEL ANALYSIS)

Loss, and Accuracy changes have been demonstrated over each epoch as displayed in Fig. 5 (a, b). When processing a complete dataset in a neural network, both feed-forward and back-propagation are used. In this case, it is crucial to comprehend the meaning of loss and accuracy as training proceeds and at which time these are constant. By understanding this scaler graph, there will be less overfitting for this reason, the above information explains and describes the scaler graph and how to avoid overfitting. By doing this, researchers observe that the increase in losses and in the accuracy of the training and validation sample go hand in hand. Based on this indicator, the proposed network in this research does not suffer from the issue of overfitting.



(a) Accuracy per Epoch **(b) Loss per Epoch**

Figure 5. CNN model analysis.

Training Results The model has been trained with 700 training images (1000 R ,1000 G, 1000 B) using the optimization method SGD with an initial learning rate of 0. 0001. To optimize the images, the dataset is passed through 10 epochs. In each epoch, the weights are adjusted in order to have an image that is closer to the desired image. Table 2 shows the training evaluation of the model's accuracy and loss, along with the associated hyperparameters. This table also shows that the number of iterations for each Epoch are 32 both the training with validation and testing images as 700, 100 and 200 images. Initially, the learning rate started with 0. 0001 and epoch by epoch it was reduced to (0. 0003).

Table 2: Training Results

		Learning rate													**Total Parameters**	
Iteration	**Batch size**	**Initial value**	**Last value**	**Loss function**	**Optimizer**	**Epochs**	**ETA**	**VAL-Loss**	**VAL-Accuracy**	**Loss**	**F1_score**	**Precision**	**Recall**	**Accuracy**	**Learning**	**Non-Learning**
32	32	0.0001	0.006	Categorical cross entropy	SGA	10	16.749 MS	0.1213	0.9880	0.0420	0.9882	0.9888	0.9876	0.9976	20,553,796	0

Testing Results As mentioned before, 100 testing images out of a total of 1000 images in COREL's image database consist of 10% of the database that has been used in the input into the model architecture and has passed through all the layers of the model. In testing, the parameters that were saved, including weights that the network reached, were utilized, and by multiplying the test images by those weights, the images were categorized into ten classes. The confusion matrix in test model proposal is presented in Fig. 6.

Figure 6. Confusion Matrix.

Here are some key metrics that can be derived from a Confusion Matrix: [24]

- Accuracy: (TP + TN) / (TP + TN + FP + FN) - This measures the overall accuracy of the model.
- Precision: TP / (TP + FP) - This measures the proportion of true positives among all predicted positives.
- Recall: TP / (TP + FN) - This measures the proportion of true positives among all actual positives.
- F1-score: 2 * (Precision * Recall) / (Precision + Recall) - This is the harmonic mean of precision and recall.
- False Positive Rate: FP / (FP + TN) - This measures the proportion of false positives among all actual negatives.
- False Negative Rate: FN / (FN + TP) - This measures the proportion of false negatives among all actual positives

Table 3: Performance of Testing.

For that can using the above rules to find the measure test for proposal model in Table 3.

Timing Test The time result of the CNN model (in seconds) for training is 16.749.234. All presented trials were conducted on a computer with a Core i7 processor, a clock rate of up to 2.7 GHz, and 16GB of memory with a preloaded Windows 10 operating system.

4.2 Phase I : Stage 2(Retrieval Module Tasks):

Loss	F1_score	Accuracy	Precision	Recall
0.0842 %	98.82%	98.52%	93.67%	91.37%

Some approaches have been suggested to improve the quality of images retrieved .In research, a proposal model to retrieve the nearest ten images from the system in Fig. 7.

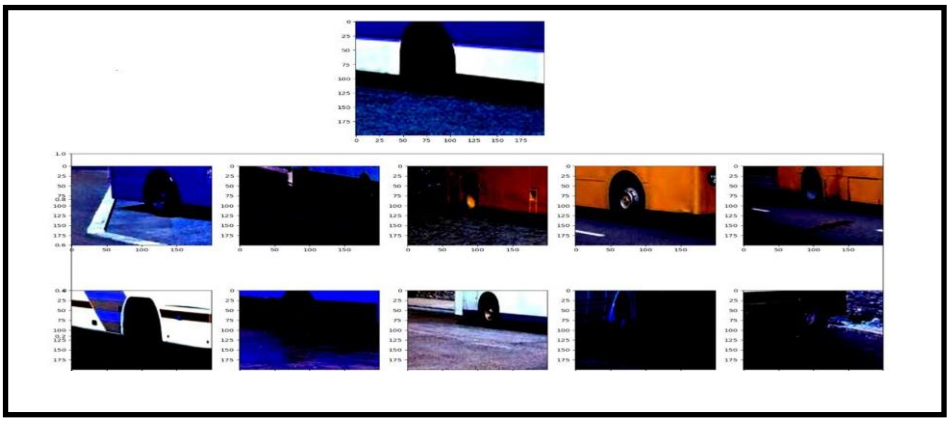


Figure 7. Ten retrieval image based on proposal model

MAP (Mean Average Precision) tests are one of the categories of evaluation metrics used for measuring the recovery effectiveness of information retrieval systems. In particular, MAP defines how well the system can retrieve the documents or other items that satisfy a given query. Table 4 shows the retrieval performance for COREL's image database in mean average precision.

Table 4. Image retrieval map for COREL's image database

MAP		
25	@5	97.4
25	@10	97.4
25	@15	96.8
25	@20	96.6
25	@25	96.4

5. Comparison with Previous Studies

Several methods are proposed in the related works to enhance the classification and retrieval images performance. Table 5 displays the summary of image classification accuracy and retrieval from various research projects.

Table 5: Image Classification Accuracy and MAP

Related works	Image classification accuracy	MAP
Kuo et al.[11]	96.9%	70%
Huang et al. [12]	90.19%	86%
Khan and Javed [13]	94.0%	91%
Manikandakumar and Karthikeyan [31]	97.79%	-
Hong et al. [32]	98.40%	-
proposed system	98.52%	97%

From this table, the researchers will hence be able to infer that the proposed method is indeed better. TensorBoard analysis of the proposed system analysis by the researcher is presented in the above section. Subsequently, select the CNN design and parameters that seem more suitable depending on the outcome of this study.

6. Conclusion

This paper has demonstrated an efficient CBIR system with the ability to semantically search the correct images based on the content with high retrieval quality. This work base on provided a proper image classification system that can semantically categorize the right images with high classification efficiency. Several algorithms were employed at once to work towards the best outcomes. In creating the deep learning model, the CNN algorithm is used to optimize the features, while the hamming distance method is used to measure the distance between the stored and input factors. It also eliminates the overfitting problem and, at the same time, enhances the model's performance by increasing the number of training images. The contribution presented is a CNN model with very high accuracy of retrieval and classification that enabled the extraction of the most significant features of the images to a level of 98.52% percent. This entails the process of escalating the number of samples of the dataset by including expected effects on images and choosing the finest features from the flatten layer, which significantly leads to high classification accuracy. These features are used in order to find the distance between each stored image feature and the input image features, which enhances the image classifying accuracy up to a certain extent and also reduces the time of image classifying. The proposed architecture in CNN is highly effective in image classification recognition tasks, as it can learn hierarchical features from the input image. CNNs are widely used in applications such as image recognition, object detection.

References

[1]C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, Rethinking the inception architecture for computer vision. 2016. doi: 10.1109/cvpr.2016.308. Available: https://doi.org/10.1109/cvpr.2016.308

[2]Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," Computer Vision and Image Understanding, vol. 192, p. 102897, Mar. 2020, doi: 10.1016/j.cviu.2019.102897. Available: https://doi.org/10.1016/j.cviu.2019.102897

[3]H. Ahn and C. Yim, "Convolutional Neural Networks Using Skip Connections with Layer Groups for Super-Resolution Image Reconstruction Based on Deep Learning," Applied Sciences, vol. 10, no. 6, p. 1959, Mar. 2020, doi: 10.3390/app10061959. Available: https://doi.org/10.3390/app10061959

[4]D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," IEEE Conference on Computer Vision and Pattern Recognition, Jun. 2012, doi: 10.1109/cvpr.2012.6248110. Available: https://doi.org/10.1109/cvpr.2012.6248110

[5]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386. Available: https://doi.org/10.1145/3065386

[6]B. Patel, K. Yadav, and D. Ghosh, “Current Trend and Methodologies of Content-Based Image Retrieval: Survey,” in Algorithms for intelligent systems, 2021, pp. 647–665. doi: 10.1007/978-981-15-6707-0_64. Available: https://doi.org/10.1007/978-981-15-6707-0_64
[7]X. Li, J. Yang, and J. Ma, “Recent developments of content-based image retrieval (CBIR),” Neurocomputing, vol. 452, pp. 675–689, Sep. 2021, doi: 10.1016/j.neucom.2020.07.139. Available: https://doi.org/10.1016/j.neucom.2020.07.139
[8]M. H. Hadid, Q. M. Hussein, Z. T. Al-Qaysi, M. A. Ahmed, and M. M. Salih, “An Overview of Content-Based Image Retrieval Methods and Techniques,” Iraqi Journal for Computer Science and Mathematics, pp. 66–78, Jul. 2023, doi: 10.52866/ijcsm.2023.02.03.006. Available: https://doi.org/10.52866/ijcsm.2023.02.03.006
[9]R. Gupta, P. Mukherjee, B. Lall, and V. Gupta, Semantics Preserving Hierarchy based Retrieval of Indian heritage monuments. 2020. doi: 10.1145/3423323.3423409. Available: https://doi.org/10.1145/3423323.3423409
[10]Z. Cao, S. Mu, Y. Xu, and M. Dong, Image retrieval method based on CNN and dimension reduction. 2018. doi: 10.1109/spac46244.2018.8965601. Available: https://doi.org/10.1109/spac46244.2018.8965601
[11]C.-H. Kuo, Y.-H. Chou, and P.-C. Chang, “Using deep convolutional neural networks for image retrieval,” Electronic Imaging, vol. 28, no. 2, pp. 1–6, Feb. 2016, doi: 10.2352/issn.2470-1173.2016.2.vipc-231. Available: https://doi.org/10.2352/issn.2470-1173.2016.2.vipc-231
[12]H.-K. Huang, C.-F. Chiu, C.-H. Kuo, Y.-C. Wu, N. N. Y. Chu, and P.-C. Chang, “Mixture of deep CNN-based ensemble model for image retrieval,” IEEE 5th Global Conference on Consumer Electronics, Oct. 2016, doi: 10.1109/gcce.2016.7800375. Available: https://doi.org/10.1109/gcce.2016.7800375
[13]U. A. Khan and A. Javed, “A hybrid CBIR system using novel local tetra angle patterns and color moment features,” Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, pp. 7856–7873, Jul. 2022, doi: 10.1016/j.jksuci.2022.07.005. Available: https://doi.org/10.1016/j.jksuci.2022.07.005
[14]K. Ma, B. Wang, Y. Li, and J. Zhang, “Image retrieval for local architectural heritage recommendation based on deep hashing,” Buildings, vol. 12, no. 6, p. 809, Jun. 2022, doi: 10.3390/buildings12060809. Available: https://doi.org/10.3390/buildings12060809
[15]L. D. Medus, M. Saban, J. V. Francés-Víllora, M. Bataller-Mompeán, and A. Rosado-Muñoz, “Hyperspectral image classification using CNN: Application to industrial food packaging,” Food Control, vol. 125, p. 107962, Feb. 2021, doi: 10.1016/j.foodcont.2021.107962. Available: https://doi.org/10.1016/j.foodcont.2021.107962
[16]S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” Mechanical Systems and Signal Processing, vol. 151, p. 107398, Nov. 2020, doi: 10.1016/j.ymssp.2020.107398. Available: https://doi.org/10.1016/j.ymssp.2020.107398
[17]X. Zhang et al., “Hierarchical bilinear convolutional neural network for image classification,” IET Computer Vision, vol. 15, no. 3, pp. 197–207, Mar. 2021, doi: 10.1049/cvi2.12023. Available: https://doi.org/10.1049/cvi2.12023
[18]A. Shabbir et al., “Satellite and scene image classification based on transfer learning and fine tuning of RESNET50,” Mathematical Problems in Engineering, vol. 2021, pp. 1–18, Jul. 2021, doi: 10.1155/2021/5843816. Available: https://doi.org/10.1155/2021/5843816
[19]J. Li and J. Z. Wang, “Real-Time computerized annotation of pictures,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 6, pp. 985–1002, Jun. 2008, doi: 10.1109/tpami.2007.70847. Available: https://doi.org/10.1109/tpami.2007.70847
[20]U. A. Khan, A. Javed, and R. Ashraf, “An effective hybrid framework for content based image retrieval (CBIR),” Multimedia Tools and Applications, vol. 80, no. 17, pp. 26911–26937, May 2021, doi: 10.1007/s11042-021-10530-x. Available: https://doi.org/10.1007/s11042-021-10530-x
[21]D. Gupta, R. Loane, S. Gayen, and D. Demner-Fushman, “Medical image retrieval via nearest neighbor search on pre-trained image features,” Knowledge-Based Systems, vol. 278, p. 110907, Aug. 2023, doi: 10.1016/j.knosys.2023.110907. Available: https://doi.org/10.1016/j.knosys.2023.110907
[22]W. Chen et al., “Deep learning for instance retrieval: a survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 6, pp. 7270–7292, Nov. 2022, doi: 10.1109/tpami.2022.3218591. Available: https://doi.org/10.1109/tpami.2022.3218591
[23]S. A. Hasan and G. M. Hadi, “Review about SIFT and Local Feature Extraction in Content Based Image Retrieval,” 5th International Conference on Communication Engineering and Computer Science, Jan. 2024, doi: 10.24086/cocos2024/paper.1533. Available: https://doi.org/10.24086/cocos2024/paper.1533
[24]E. M. Alsaedi and A. K. Farhan, “Retrieving encrypted images using convolution neural network and fully homomorphic encryption,” Baghdad Science Journal, vol. 20, no. 1, p. 0206, Feb. 2023, doi: 10.21123/bsj.2022.6550. Available: https://doi.org/10.21123/bsj.2022.6550
[25]Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, “Recent advances in convolutional neural network acceleration,” Neurocomputing, vol. 323, pp. 37–51, Jan. 2019, doi: 10.1016/j.neucom.2018.09.038. Available: https://doi.org/10.1016/j.neucom.2018.09.038

[26]C. Lu, M. Kozakai, and L. Jing, “Sign Language Recognition with Multimodal Sensors and Deep Learning Methods,” Electronics, vol. 12, no. 23, p. 4827, Nov. 2023, doi: 10.3390/electronics12234827. Available: https://doi.org/10.3390/electronics12234827
[27]R. Raj and A. Kos, “An improved human activity recognition technique based on convolutional neural network,” Scientific Reports, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49739-1. Available: https://doi.org/10.1038/s41598-023-49739-1
[28]J. Cheng, M. Sadiq, O. A. Kalugina, S. A. Nafees, and Q. Umer, “Convolutional neural network based approval prediction of enhancement reports,” IEEE Access, vol. 9, pp. 122412–122424, Jan. 2021, doi: 10.1109/access.2021.3108624. Available: https://doi.org/10.1109/access.2021.3108624
[29]Y. H. Ali et al., “Optimization system based on convolutional neural network and internet of medical things for early diagnosis of lung cancer,” Bioengineering, vol. 10, no. 3, p. 320, Mar. 2023, doi: 10.3390/bioengineering10030320. Available: https://doi.org/10.3390/bioengineering10030320
[30]A. F. Al-Zubidi, A. K. Farhan, and S. M. Towfek, “Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model,” Journal of Intelligent Systems, vol. 33, no. 1, Jan. 2024, doi: 10.1515/jisys-2023-0195. Available: https://doi.org/10.1515/jisys-2023-0195
[31]M. Manikandakumar and P. Karthikeyan, “Weed classification using particle swarm optimization and deep learning models,” Computer Systems Science and Engineering, vol. 44, no. 1, pp. 913–927, Jan. 2023, doi: 10.32604/csse.2023.025434. Available: https://doi.org/10.32604/csse.2023.025434
[32]M. Hong, B. Rim, H. Lee, H. Jang, J. Oh, and S. Choi, “Multi-Class Classification of lung diseases using CNN models,” Applied Sciences, vol. 11, no. 19, p. 9289, Oct. 2021, doi: 10.3390/app11199289. Available: https://doi.org/10.3390/app11199289