

Pre-trained CNNs Models for Content based Image Retrieval

Ali Ahmed

Faculty of Computing and Information Technology
King Abdulaziz University –Rabigh
Rabigh 21589, Saudi Arabia

Abstract—Content based image retrieval (CBIR) systems is a common recent method for image retrieval and is based mainly on two pillars extracted features and similarity measures. Low level image presentations, based on colour, texture and shape properties are the most common feature extraction methods used by traditional CBIR systems. Since these traditional handcrafted features require good prior domain knowledge, inaccurate features used for this type of CBIR systems may widen the semantic gap and could lead to very poor performance retrieval results. Hence, features extraction methods, which are independent of domain knowledge and have automatic learning capabilities from input image are highly useful. Recently, pre-trained deep convolution neural networks (CNN) with transfer learning facilities have ability to generate and extract accurate and expressive features from image data. Unlike other types of deep CNN models which require huge amount of data and massive processing time for training purposes, the pre-trained CNN models have already trained for thousands of classes of large-scale data, including huge images and their information could be easily used and transferred. ResNet18 and SqueezeNet are successful and effective examples of pre-trained CNN models used recently in many machine learning applications, such as classification, clustering and object recognition. In this study, we have developed CBIR systems based on features extracted using ResNet18 and SqueezeNet pre-trained CNN models. Here, we have utilized these pre-trained CNN models to extract two groups of features that are stored separately and then later are used for online image searching and retrieval. Experimental results on two popular image datasets Core-1K and GHIM-10K show that ResNet18 features based on the CBIR method have overall accuracy of 95.5% and 93.9% for the two datasets, respectively, which greatly outperformed the traditional handcraft features based on the CBIR method.

Keywords—Pre-trained deep neural networks; transfer learning; content based image retrieval

I. INTRODUCTION

The great development of digital computers and various smart devices, in addition to the large and steady increase in the different storage media, led to a considerable increase in digital images and other types of multimedia components. The large amount of multimedia, especially digital images, are used in many fields of medical treatment, satellite data and remote sensing, digital forensics and digital evidence [1, 2]. The large and rapid increase of the size of the digital content of images relies basically on retrieving these images from their various sources so that they can be

used in the specific field or application. The content-based image retrieval (CBIR) method is one of the modern and effective ways to retrieve images from various image repositories, as well as from the web. CBIR is defined as the process of image retrieval by extracting some useful information from their low-level features or contents, such as colour, texture and shape, or other level of characteristics. The efficiency and effectiveness of any content-based image retrieval (CBIR) system depends on the extracted features because it will be used as numerical values in calculating similarity between the query submitted by the end user and all the images stored in a repositories or data storage [3]. One of the main challenges facing any content-based image retrieval (CBIR) is the semantics gap, which is defined as missing or lost information as a result of representing or capturing an image using an imaging device and the human vision system (HVS) used to perceive that image. This semantics gap that exists between the visual information captured by the imaging device and HVS can be reduced either by including domain or field-specific knowledge or by using some machine learning techniques to be trained and act like HVS. There has been a great development in the last decades in machine learning techniques and methods; these techniques have proven successful in being used in many areas of application, such as classification, clustering as well as information retrieval. Many machine learning methods have achieved great success and good results in many studies related to image retrieval. The main reason for that success is the availability of large amounts of images and pre-classified data in addition to the high computing capabilities of modern computers. The convolutional neural network (CNN) is a group of nonlinear transforming processes that have the ability to learn from the input data. These networks learn different features, especially image features. It uses small squares of the input data, and then it applies a set of operations of filter scanning for the input pixel values, known as convolutional operations. Deep convolutional neural networks are used in many applications related to digital image processing, such as image clustering, image classification and pattern or object recognition. On the other hand, these convolutional neural networks require huge data, computational resources and processing time. Pre-trained deep learning neural networks are the latest developed methods of convolutional neural networks that have been applied recently and have demonstrated high accuracy and good results in many areas of research. The superior ability of pre-trained networks are the

result of their training on large-scale images for a large number of classes. This facility enables users to benefit from the advantage of pre-training and the transfer learning concept in various processes of classification or feature extraction. These pre-trained CNN models, such as AlexNet, GoogleNet, SqueezeNet and ResNet-18, have been applied for solving many problems, such as pattern recognition, computer vision, natural language processing, and medical image classification. Due to the success and good performance of this type of neural network, in this study we propose the CBIR method that is based on the two popular types of these networks in extracting features, which will be used to retrieve images through their content. The remainder of this study is organized as follows: Section 2 refers to related studies and state-of-art methods and approaches used in the area of CBIR, while Section 3 presents and explains our proposed methodology in more detail. Results and findings are reported and discussed in Section 4, while in Section 5 we summarize and conclude our study.

II. RELATED WORK

Image retrieval is an old research problem of which the idea is to retrieve images like the user's query from the image data repository. The traditional method used for this process, which is known as text based image retrieval (TBIR) has used keywords associated with each image. These keywords are designed and indexed manually and later used to search for similar images. There are many drawbacks in this method, including the human effort to design the keywords, cost and time, which is extremely labour intensive, in addition to the low accuracy and efficiency of the retrieval process because it relies on searching for index words and not the content of the images. Moreover, this traditional method does not enable the developers to describe the meanings and semantics of images content in databases, especially those that contain a large number of images. All of the previous limitations led to dispensing with this old method and replacing it with the new and modern method, which is known as content-based image retrieval (CBIR) [4, 5]. In this modern method of image retrieval, low level image representations are used in the comparison or similarity process. These representational properties or features are extracted directly from the images. In most content-based image retrieval (CBIR) systems, the lowest content representations of the query image are compared against the representations of all the images in the database, and then the most similar images are retrieved. The most common visual properties used in this method are the characteristics of colour, texture, and shape [6-8]. The colour feature is one of the most used characteristics in retrieving images in CBIR. The colour is one of the best distinguishing features at lower-level visual features of CBIR. The colour is also one of the effective, robust, and easy-to-implement properties and requires less storage capacity [9]. The histogram is one of the best methods of representing colours that can be used in CBIR [10]. The researchers in [11] have designed the CBIR systems using the global colour histogram method and they achieved acceptable results. The Hue Saturation Value (HSV) colour space representation is useful and has better retrieval results compared with the default Red Green Blue (RGB)

colour space [12]. The texture descriptor is the second most widely used features representation space in CBIR. The gray-level co-occurrence matrix (GLCM) [13] is the popular method used to extract many useful texture features, such as uniformity, correlation, contrast and entropy [14]. Features learning algorithms based on convolutional neural networks (CNNs) have become popular and widely used in recent years, due to their capabilities and powerful in many disciplines in general and in image processes in particular due to their accuracy and good performance in retrieval tasks [15]. One major drawback that faces this type of new generation of neural networks is their need of huge training data, which is not available across various domains of knowledge [16]. Due to this training limitation, the latest generation of pre-training CNN is found to fit the needs for well-trained CNNs and has highly accurate results. These pre-training CNN models have the ability to transfer their knowledge because they were trained on big-scale annotated natural image data collections in ImageNet [17] and it was successfully applied in many image processing application area [18-22]. There are three different methods that could be used to obtain the benefits and utilized the power of these pre-trained CNN and transfer their learning capabilities. These methods are feature extraction, using their architectures with proper needed tuning and, lastly, we can train some layers of the model while freezing others. In this study we utilize the pre-trained CNN and propose the content-based image retrieval method based on the features extracted using their pre-trained architecture. The contributions of this study can be summarized as follows:

- To utilize the ResNet-18 and SqueezeNet pre-trained CNN model for feature extraction from images collections.
- To develop a retrieval method based on extracted features and Euclidean distance similarity measures for CBIR.
- To enhance the retrieval process and compare the performance of our proposed method with some other state-of-the-art methods.

III. METHODOLOGY

The proposed method of this study consists of two phases or processes – offline process and online process – as shown in Fig. 1. In the offline process, the pre-trained CNN model is used for feature extraction while the online phase is responsible for end user query manipulation and retrieval results. The pre-trained deep CNN models consist of many layers that apply their learning process in an incremental manner and execute many subsampling and convolutions process. Here, SqueezeNet and ResNet18 pre-trained deep CNN models are used for feature extraction, and the features vector is saved in the features database to be used later for a similarity calculation. The online phase is the most important phase in which the extracted features generated in the previous step is used instead of the image itself for matching and similarity computation, and then the top similar images are retrieved to the end user.

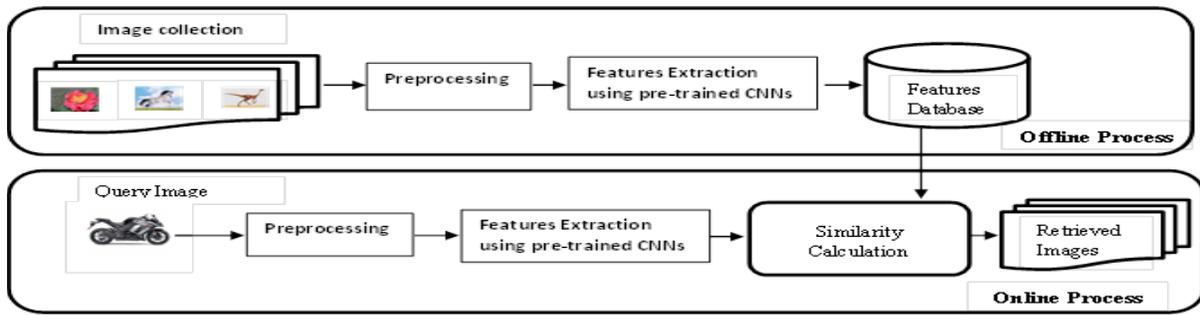


Fig. 1. Main Framework of CBIR.

A. SqueezeNet Pre-trained CNNs for Features Extraction

SqueezeNet is a simple and effective pre-trained CNN architecture with acceptable performance, and it has had successful usage recently. This model consists of 68 layers and it requires a 227x227x3 size input image [23]. After each image is set to the required size, the process in this model goes through 14 convolutions processing block elements with different rescaling and resampling operations, as shown in Fig. 2. Finally, a total of 1000 features are extracted and saved into a separated database for future use by the next online similarity and ranking process.

B. ResNet18 Pre-trained CNNs for Features Extraction

The second pre-trained CNN model used here is ResNet-18, a convolutional neural network that consists of 18 layers deep that was developed by [24]. Both this model and the previous model are trained on more than a million images from the ImageNet database [17]. This wide-range training process is very important for the transfer learning process as we mentioned earlier. A total of 512 features are extracted from the last fully connected layer, as shown in Fig. 3. The number of convolutional blocks and size of each block are also shown in Fig. 3. The group of features extracted here is saved again for a further similarity calculation process.

C. Visual Features based on Color and Texture Descriptors

In this method, a total of 18 colour features are extracted from each image using six colour moments. Each colour image is converted from the RGB colour space to HSV colour representation, and then six features are extracted from each channel. For the texture descriptor, four functions are used to extract texture features using the gray-level co-occurrence matrix (GLCM). Again, by using three channels of HSV, a total of 12 features are combined with the previous 18 colour features into a single vector of 30 features. The retrieval performance of these features is used as a based result to compare with our proposed method results and findings. The finding and performance of the features extracted by the previous two pre-trained CNN models and this group of traditional features are analyzed and compared at the next section.

D. Similarity Measure

The group features are extracted from both pre-trained deep CNN models, and traditional colour and texture features vectors are stored in separated databases to perform the similarity measures, which is considered an important

online phase for the retrieval process. For this purpose, this study uses Euclidean distance that is considered the standard similarity coefficient used by many related studies [25]. For the two images vectors X and Y of numeric values, the similarity measure is calculated using the following equation.

$$[\sum_i^n (x_i - y_i)^2]^{1/2}$$

where n is the number of dimensions of the X and Y vectors.

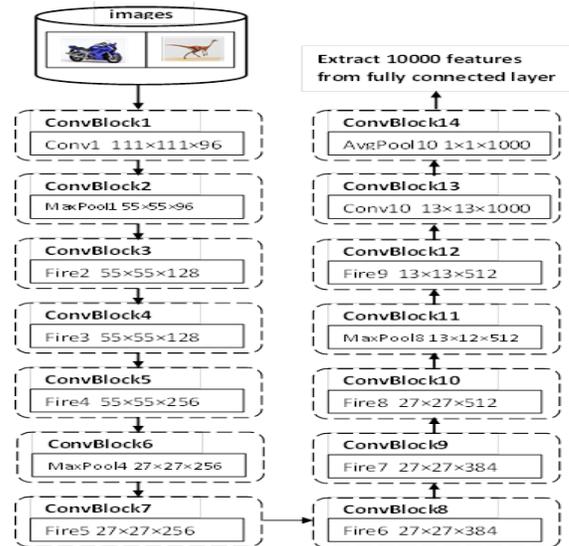


Fig. 2. SqueezeNet Model for Feature Extraction.

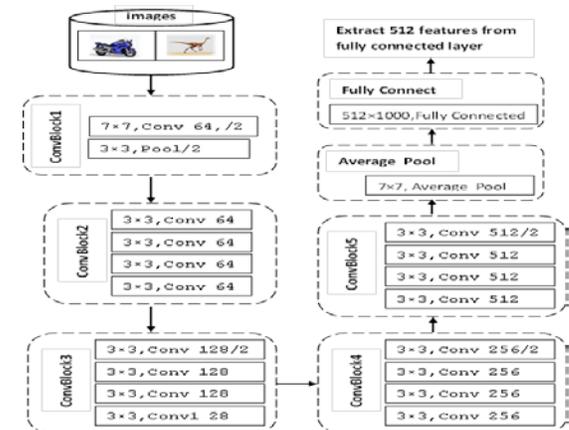


Fig. 3. ResNet18 Model for Feature Extraction.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. Images Datasets

For the experiments and evaluation of our proposed method, the study uses two of the most known image datasets. These two datasets were used widely in many studies related to CBIR and their recent usage found in [26]. First dataset is Corel-1K [27] and has a total of 1000 images divided into ten categories with 100 images for each class. The resolution of images is 256×384 or 384×256 pixels. GHIM-10K [28] is the second dataset used here; it consists of 10000 images divided equally into 20 classes with 500 images with a resolution of 300×400 pixels or 400×300 pixels. The second image dataset is 10 times larger, and it has more challenges than the previous one, since it contains more classes with a larger number of images. Samples from both datasets are shown in Fig. 4 and Fig. 5, where a single image from each class has been taken.

B. Performance Evaluation

Recall and precision are two performance measures used here. These two metrics could be used for evaluation of any retrieval model, especially the CBIR model. The performance based on these metrics could compute at any point of the top retrieved image, but for simplicity many researchers calculated their results at the top ten retrieved images. The general formula for these metrics is shown in the following equations.

$$\text{Recall} = \frac{\text{number of relevant images retrieved}}{\text{total number of retrieved images}}$$

$$\text{Precision} = \frac{\text{number of relevant images retrieved}}{\text{total number of relevant images}}$$

C. Results and Discussion

In this study, three different retrieval methods are developed and their results and findings are evaluated. The result of each experiment is reported; tables and figures of their findings are illustrated, as we will show in the coming paragraphs. The first retrieval method is based on traditional colour and texture features, which is considered the based model before the revolution of deep convolution models. These traditional image descriptors are wide image descriptors that have been used for CBIR for many years and have good performance results if followed and combined with some enhancement techniques such as relevance feedback or expansion processes. The second and third retrieval methods are based on SqueezeNet and ResNet18 features, respectively. For all our experiments, a random 10 images from each image class are selected for user queries and then average recall and precision at the top 10 retrieved images are calculated. The overall results of the Corel-1K images database for the three retrieval methods are shown in Table I. CL and GLCM refer to the first method since we use the colour moment functions for colour features and the GLCM method for texture features spaces. In this table, recall and precision for each of the 10 classes as well as the average values for the two metrics are shown. A fast inspection of these values illustrated that the two pre-trained CNN models

outperformed the traditional colour and texture based retrieval method. Bolded cell values in both tables are used to represent the highest value of average precision of each class among the three retrieval methods, proving that ResNet18 has better retrieval results compared with SqueezeNet, as well as the traditional feature based retrieval method. For the GHIM-10K images database, which is considered more challenging; our two pre-trained CNN models also have better retrieval performance. For all 20 classes, ResNet18 has many highest retrieval values for xx classes out of 20 classes and moreover, the overall average recall and precision outperformed the SqueezeNet and the traditional method. The final result for this images database is shown in Table II. More analytical results are shown in Fig. 6 and Fig. 7 for both image databases. These two figures related images are retrieved in different top values (from 5 to 100) with precision values. The higher location of ResNet18 plotted lines proved that the good performance for these features outperformed the based model. Finally, visual retrieval performance in terms of top image retrieved for each query image for some selected classes for both images databases, show that ResNet18 has the best achievement. It was success to retrieve all correct images from the retrieved top 10 images as shown in Fig. 8. Part (a) of this figure represents the top retrieved images for the Bus class of the GHIM-10K dataset. Seven out of ten images are successfully retrieved using CL and GLCM features, compared with 10 out of 10 images retrieved using both ResNet18 and the SqueezeNet based model, as shown in part (b). Another example is shown in part (c), (d) and (e) for the Bikes class, 7 out of 10 and 9 out of 10 d are retrieved for GLCM and SqueezeNet, while all 10 images are retrieved using the ResNet18 model. Part (f) of Fig. 8 also represents the top retrieved images of Core-1K images. For the Dinosaurs class, our proposed pre-trained models and CL and GLCM traditional model have successfully retrieved all top 10 images. This is due to the simplicity of images of this class as its colour and texture properties are very clear and simple.



Fig. 4. Samples Images from Corel-1K Dataset.



Fig. 5. Samples Images from GHIM-10K Dataset.

TABLE I. AVERAGE RECALL AND PRECISION FOR COREL-1K DATASET

Class	CL and GLCM		SqueezeNet Features		ResNet18 Features	
	R	P	R	P	R	P
Africa	0.038	0.38	0.087	0.87	0.097	0.97
Beach	0.037	0.37	0.056	0.56	0.077	0.77
Building	0.051	0.51	0.077	0.77	0.089	0.89
Buses	0.072	0.72	0.1	1	0.1	1
Dinosaurs	0.1	1	0.1	1	0.1	1
Elephants	0.042	0.42	0.1	1	0.1	1
Flowers	0.086	0.86	0.1	1	0.1	1
Horses	0.071	0.71	0.096	0.96	0.099	0.99
Mountains	0.037	0.37	0.087	0.87	0.098	0.98
Food	0.029	0.29	0.091	0.91	0.095	0.95
Mean	0.0563	0.5630	0.0894	0.8940	0.0955	0.9550

TABLE II. AVERAGE RECALL AND PRECISION FOR GHIM-10K DATASET

Class	CL and GLCM		SqueezeNet Features		ResNet18 Features	
	R	P	R	P	R	P
Fireworks	0.0124	0.62	0.0188	0.94	0.02	1
Buildings	0.0064	0.32	0.0152	0.76	0.0196	0.98
Walls	0.0052	0.26	0.015	0.75	0.0182	0.91
Cars	0.0052	0.26	0.02	1	0.02	1
Flies	0.0074	0.37	0.016	0.8	0.0198	0.99
Mountains	0.0042	0.21	0.0134	0.67	0.0188	0.94
Flowers	0.0072	0.36	0.019	0.95	0.02	1
Trees	0.0098	0.49	0.018	0.9	0.0192	0.96
Green Grounds	0.0126	0.63	0.0178	0.89	0.0176	0.88
Beaches	0.0052	0.26	0.0142	0.71	0.0174	0.87
Aeroplanes	0.0074	0.37	0.0176	0.88	0.0198	0.99
Butterflies	0.007	0.35	0.0172	0.86	0.019	0.95
Forts	0.0058	0.29	0.0144	0.72	0.0156	0.78
Sunsets	0.0138	0.69	0.0186	0.93	0.02	1
Bikes	0.0064	0.32	0.0194	0.97	0.02	1
Boats	0.0058	0.29	0.02	1	0.02	1
Ships	0.0048	0.24	0.0162	0.81	0.0184	0.92
Chickens	0.0074	0.37	0.0188	0.94	0.02	1
Insects	0.0068	0.34	0.014	0.7	0.0138	0.69
Horses	0.0062	0.31	0.0176	0.88	0.0184	0.92
Mean	0.0074	0.3675	0.0171	0.8530	0.0188	0.9390

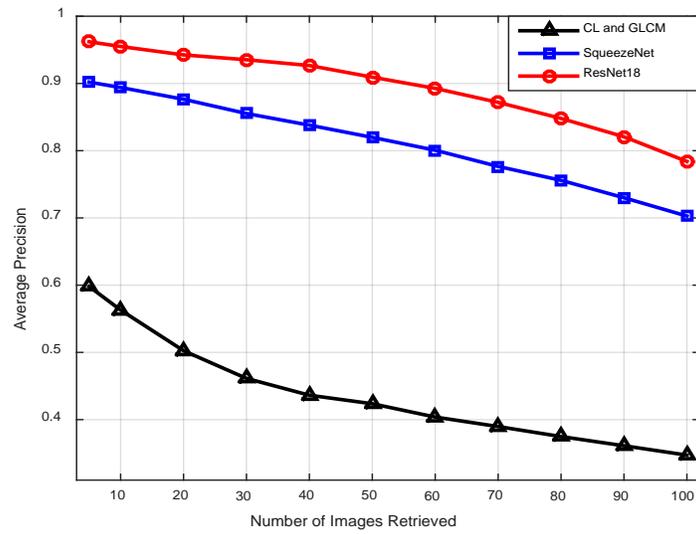


Fig. 6. Average Precision vs Number of Images Retrieved for Corel-1K

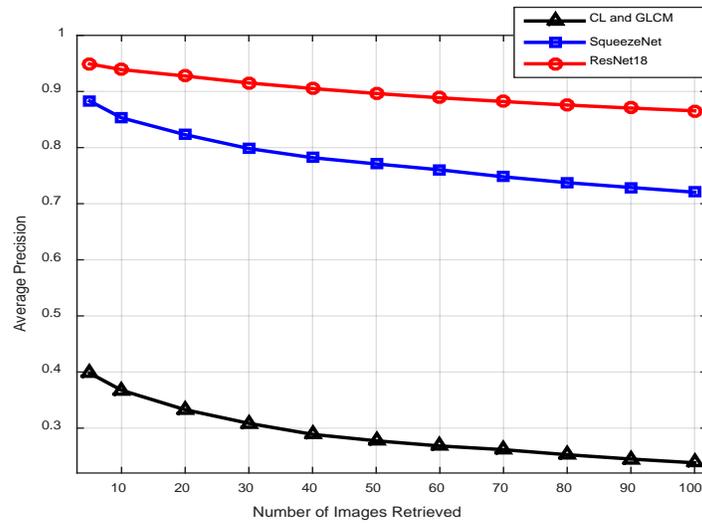


Fig. 7. Average Precision vs Number of Images Retrieved for GHIM-10K.

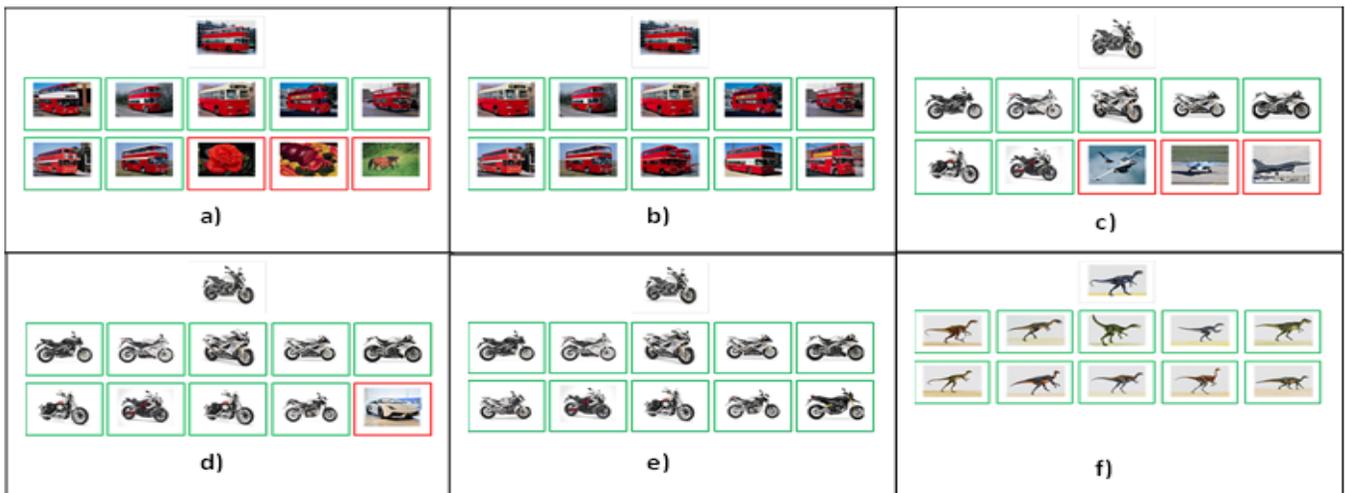


Fig. 8. Samples of Top Retrieved Images for some Classes from Corel-1k and GHIM-10K Datasets.

From visual inspection as shown in Fig. 8, the good retrieval performance of pre-trained CNN models for these most challenging GHIM-10K dataset images is clearly proved. There are similar best results in terms of top retrieved images found in other classes for both datasets such as Elephants and Flowers classes of Corel-1K and Sunsets and Boats of GHIM-10K dataset images.

V. CONCLUSION

In this study, we implemented and tested the CBIR method based on features extracted from two different pre-trained deep CNN models. CBIR models based on SqueezeNet and ResNet18 models are developed and tested on multi-class digital images datasets (Core-1K and GHIM-10K) and their results were compared with traditional CBIR, based on colour and texture features descriptors. The study achieved the average retrieval precisions of 89.40% and 95.50% for Corel-1K and 85.30% and 93.90% for GHIM-10K, using SqueezeNet and ResNet18 respectively, which clearly outperformed the CBIR model based on colour and texture-based features. Our average precisions of the ResNet18 based CBIR method are increased by 39.20% and 56.55%, compared with colour and texture based CBIR for Core-1K and GHIM-10K, respectively, which was clearly considered to be effective and has the best retrieval performance. Our retrieval performance results of ResNet18 have better performance compared with the SqueezeNet based CBIR method, and this could be due to accurate extracted features, compared with the large number of features extracted by the SqueezeNet pre-trained CNN model. For future efforts, other popular pre-trained CNN models could be used for feature extraction purposes, which could achieve better performance after proper and required tuning processes for their architectures.

REFERENCES

- [1] Lin, X., J.-H. Li, S.-L. Wang, F. Cheng and X.-S. Huang, Recent advances in passive digital image security forensics: A brief review. *Engineering*, 2018. 4(1): p. 29-39.
- [2] Fonseca, L.M.G., L.M. Namikawa, and E.F. Castejon., Digital image processing in remote sensing. in 2009 Tutorials of the XXII Brazilian Symposium on Computer Graphics and Image Processing. 2009. IEEE.
- [3] Liu, Y., D. Zhang, G. Lu and W.-Y. Ma., A survey of content-based image retrieval with high-level semantics. *Pattern recognition*, 2007. 40(1): p. 262-282.
- [4] Zare, M.R., R.N. Aïnon, and W.C. Seng., Content-based image retrieval for blood cells. in 2009 Third Asia International Conference on Modelling & Simulation. 2009. IEEE.
- [5] Sreedevi, S. and S. Sebastian., Content based image retrieval based on Database revision. in 2012 International Conference on Machine Vision and Image Processing (MVIP). 2012. IEEE.
- [6] Latif, A., A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratal, B. Zafar, S. H. Dar, M. Sajid and T. Khalil., Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, 2019. 2019.
- [7] Wang, S., K. Han, and J. Jin., Review of image low-level feature extraction methods for content-based image retrieval. *Sensor Review*, 2019.
- [8] Thilagam, M. and K. Arunish., Content-Based Image Retrieval Techniques: A Review. in 2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW). 2018. IEEE.
- [9] Chen, Y., The image retrieval algorithm based on color feature. in 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). 2016. IEEE.
- [10] Zhang, Z., W. Li, and B. Li., An improving technique of color histogram in segmentation-based image retrieval. in 2009 Fifth International Conference on Information Assurance and Security. 2009. IEEE.
- [11] Chmelar, P. and A. Benkril. Efficiency of HSV over RGB Gaussian Mixture Model for fire detection. in 2014 24th International Conference Radioelektronika. 2014. IEEE.
- [12] Nazir, A., R. Ashraf, T. Hamdani and N. Ali., Content based image retrieval system by using HSV color histogram, discrete wavelet transform and edge histogram descriptor. in 2018 international conference on computing, mathematics and engineering technologies (iCoMET). 2018. IEEE.
- [13] Weszka, J.S., C.R. Dyer, and A. Rosenfeld., A comparative study of texture measures for terrain classification. *IEEE transactions on Systems, Man, and Cybernetics*, 1976(4): p. 269-285.
- [14] Ahmed, A. and S. Mohamed, Implementation of early and late fusion methods for content-based image retrieval. *International Journal of Advanced and Applied Sciences*, 8(7) 2021, Pages: 97-105.
- [15] Rezazadeh, N., Initialization of weights in deep belief neural network based on standard deviation of feature values in training data vectors. *Vol (6)*, 2017(6): p. 708-715.
- [16] Han, D., Q. Liu, and W. Fan, A new image classification method using CNN transfer learning and web data augmentation. *Expert Systems with Applications*, 2018. 95: p. 43-56.
- [17] ImageNet. <http://www.image-net.org>.
- [18] Lou, Y., G. Fu, Z. Jiang, A. Men and Y. Zhou, PT-NET: Improve object and face detection via a pre-trained CNN model. in 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP). 2017. IEEE.
- [19] Rajesh, K.B., Cybernetic microbial detection system using transfer learning. *Multimedia Tools and Applications*, 2018: p. 1-18.
- [20] Marmanis, D., M. Datcu, T. Esch and U. Stilla., Deep learning earth observation classification using ImageNet pretrained networks. *IEEE Geoscience and Remote Sensing Letters*, 2015. 13(1): p. 105-109.
- [21] Hussain, M., J.J. Bird, and D.R. Faria. A study on cnn transfer learning for image classification. in UK Workshop on computational Intelligence. 2018. Springer.
- [22] Akcay, S., M. E. Kundegorski, C. G. Willcocks and T. P. Breckon., Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE transactions on information forensics and security*, 2018. 13(9): p. 2203-2215.
- [23] Gopalakrishnan, K., S. K. Khaitan, A. Choudhary and A. Agrawal., Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection. *Construction and Building Materials*, 2017. 157: p. 322-330.
- [24] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
- [25] Shirshorshidi, A.S., S. Aghabozorgi, and T.Y. Wah, A comparison study on similarity and dissimilarity measures in clustering continuous data. *PLoS one*, 2015. 10(12).
- [26] Varish, N., A. K. Pal, R. Hassan, M. K. Hasan, A. Khan, N. Parveen, D. Banerjee, V. Pellakuri, A. U. Haqis and I. Memon., Image retrieval scheme using quantized bins of color image components and adaptive tetrolet transform. *IEEE Access*, 2020. 8: p. 117639-117665.
- [27] Li, J. and J.Z. Wang, Real-time computerized annotation of pictures. *IEEE transactions on pattern analysis and machine intelligence*, 2008. 30(6): p. 985-1002.
- [28] Liu, G.-H., J.-Y. Yang, and Z. Li, Content-based image retrieval using computational visual attention model. *pattern recognition*, 2015. 48(8): p. 2554-2566.