Hybrid Shallow Learning and Deep Learning for Feature Extraction and Image Retrieval

Hanen Karamti[1,2][a], Hadil Shaiba[1][b] and Abeer M.Mahmoud[1,3][c]
[1]Computer Sciences Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, PO Box 84428, Riyadh, Saudi Arabia
[2]MIRACL Laboratory, ISIMS, University of Sfax, B.P. 242, 3021 Sakiet Ezzit, Sfax, Tunisia
[3]Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt

Keywords: Deep Convolutional Neural Network, Image Retrieval, Fuzzy C-Means Clustering, Feature Extraction.

Abstract: In the last decennia, several works have been developed to extract global/or local features from images. However, the performance of image retrieval stay surfing from the problem of semantic interpretation of the visual content of images (semantic gap). Recently, deep neural networks (DCNNs) showed excellent performance in different fields like image retrieval for feature extraction compared to traditional techniques. Although, Fuzzy C-Means (FCM) Clustering Algorithm that is a shallow learning method, but it has a competitive performance in the clustering field. In this paper, we present a new method for feature extraction combining DCNN and Fuzzy c-means, where DCNN gives a compact representation of images and FCM clusters the features and enhances the real-time for searching. The proposed method is performed against other methods in literature on two benchmark datasets: Oxford5K and Inria Holidays, where the proposed method overbeats respectively 83% and 86%.

1 INTRODUCTION

Every day, many numerical images are born in several models like medicine, science, and biology. This big mass needs efficient tools for indexing, searching and retrieving images like the CBIR (Content-Based Image retrieval) systems. CBIR has been a hot research topic in computer vision. It searches images from datasets based on their visual content. Visual content means low-level features, called also visual features, extracted from images by local and/or global descriptors. Global descriptors (Varish et al., 2015, Nazir et al., 2018) extract from images color, texture and Shape (Yuan et al., 2011, Hiremath et al., 2007, Yu et al. 2013). Color (Gopa et al., 2015) represents the most popular unit to describe the visual content of images. Color histograms, Color layout, and scalable color are the most used descriptors. The texture is the homogeneity surface quality of the object. To extract textures from images, several descriptors are proposed like the statistical methods, spectral (Qin et al., 2015).

The shape (Wu et al., 2009) is the two-dimensional object contoured by a line. The basic geometric types of shapes are oval, square, triangle circle and rectangle. Several shape descriptors are proposed like Fourier descriptors, moment invariants (Varish et al., 2015, Hiremath et al., 2007).

Local descriptors (Wang et al., 2007) are based on the low-level description of images but focused on a specific area or region in it. The most popular descriptors in this category are FAST (Features from accelerated segment test), Harris corner detectors, points of interest (POI) detectors, Scale-invariant feature transform (SIFT) (Sun et al., 2017, Yuan et al., 2011), blob detectors, Speeded Up Robust Features (SURF) (Sun et al., 2017) and Locally Aggregated Descriptors (VLAD) (Jégou et al., 2010). To calculate the similarity between images, we compare their low-level features. Several measures

[a] https://orcid.org/0000-0001-5162-2692
[b] https://orcid.org/0000-0003-1652-6579
[c] https://orcid.org/0000-0002-0362-0059

Karamti, H., Shaiba, H. and Mahmoud, A.
Hybrid Shallow Learning and Deep Learning for Feature Extraction and Image Retrieval.
DOI: 10.5220/0009417501650172
In *Proceedings of the 22nd International Conference on Enterprise Information Systems (ICEIS 2020) - Volume 1*, pages 165-172
ISBN: 978-989-758-423-7

are used like Euclidian, Manhattan, and Chi2 (Zhou t al, 2019).

Many CBIR systems (Paulin et al., 2015, Piras et al., 2010) were developed to extract the visual features from images. The traditional techniques were developed around the use of global or local descriptors and the use of similarity measures. Other systems were developed to enhance retrieval performance by the use of machine learning techniques likes SVM (Fu et al., 2016), k-means (Karamti et al., 2013), neural network (Karamti et al., 2018), etc.

Recently, deep learning (Nguyen et al., 2018, Gong et al., 2014, Qassim et al., 2018, Howard et al., 2017, Vincent et al., 2010) achieves an important success in different domains, especially in image retrieval. Therefore, this new field encourages many researchers to integrate it into their researches to enhance clustering, recognition and classification tasks.

In (Razavian et al., 2014), they developed three models of CNN for feature extraction: basic CNN, CNN with a linear SVM and CNN with additional information including the cropped and rotated examples. Radenovic et al. (Radenovic et al., 2017) proposed a new model for image retrieval called Generalized-Mean (GeM) that learned a 3D model and employed the PCA to improve the vector of features. Authors in (Mohedano et al., 2016), built a simple pipeline for retrieving that employs CNN with encoded convolutional features combined with the bag of words (BoW). Arandjelovic et al. (Arandjelovic et al., 2016) developed a new method NetVLAD that employs a new CNN architecture based on VLAD for image recognition.

In (Gordo et al., 2016), authors represented a CNN model fully connected that uses many convolutional layers combined with global descriptors after features reduction. Authors in (Qin et al., 2015) added a VLAD in CNN in the last layer of model, which performed excellent results in image retrieval. In (Qin et al., 2015, Razavian et al., 2016) authors proposed other methods for deep learning improvement as the feature extraction using Neural codes (Babenko et al., 2014) or Spatial pooling (Razavian et al., 2016).

In (Fu et al., 2016), the authors used a hybrid model based on CNN and a support vector machine (SVM). In (Ghrabat et al., 2019), authors extracted global features like texture (co-occurrence matrix) and they clustered features using the technique. Then, they applied the GA (genetic algorithm) to classify features by the use of SVM. The proposed deep model is CNN.

Recently, several deep learning techniques are proposed but DCNN was recommended as the best enhancement in image retrieval (Babenko et al., 2015, Kim et al., 2018, Lin et al., 2018). This enhancement concerns the reduction of computational cost, which is the target of many researchers. DCNN learns several features automatically with consideration of the semantic gap factor (Babenko et al., 2015, Kim et al., 2018, Lin et al., 2018, Gordo et al., 2016). In addition, deep learning mechanism can handle a large and different collection of images that represents a good performance in several tasks of image retrieval (He, et al., 2014). In addition, DCNN (Wu et al., 2018, Wu et al., 2018) was used for text classification, and the obtained results are performant again other methods that use the shallow techniques for classification.

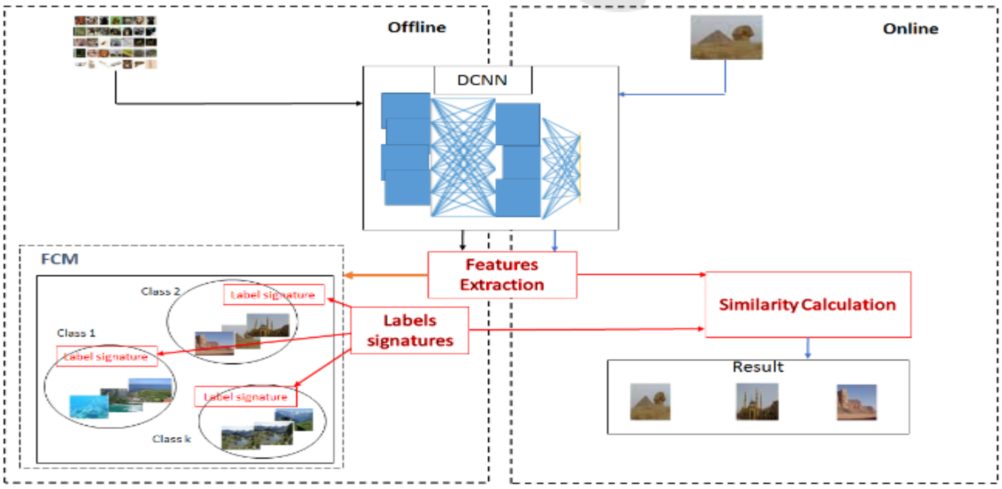

Figure 1: Feature extraction and retrieving process.

In this paper, we propose a new CBIR system that profits from the DCNN and FCM advantages. A set of experiments were presented to test the proposed method using two benchmarks: Oxford5K and Inria Holidays, where obtained results displayed promising results. The remaining paper is organized as follows: section 2 describes the proposed method. Section 3 discusses the results. Finally, section 4 presents the conclusion and further research directions.

2 CONTENT-BASED IMAGE RETRIEVAL USING DCNN AND FUZZY C-MEANS

In this paper, we profit from the benefits of DCNN and FCM to minimize the sharing time. The use of DCNN replace the visual descriptors to extract the low-level features. This convolutional model is a deep representation, composed by a set of nonlinear operations belonging to multiple levels. FCM is used to enhance accelerate the retrieving process.

Our approach, represented by Figure 1, contains three main phases: the first phase concerns features extraction. Second phase is clustering and the third phase is similarity calculation. The first phase takes place offline where features are extracted from images using the DCNN model. Then, the FCM technique is applied in the second phase to cluster images into similar clusters. After that, a label signature is assigned to each cluster to distinguish between clusters. All the labels reconstruct the output layer of DCNN that serves in retrieving phase. Last phase represents the retrieving phase that calculates similarity between the features extracted from query and label signatures. Based on the similarity values, the closest class is returned containing the relevant images.

2.1 Deep Neural Network

To extract the low-level features from images, we adopted the Deep Convolutional Neural Network architecture (DCNN). Several researches in the image retrieval field focused on the use of DCNN to extract the features because this algorithm is able to produce better results despite its high computational cost.

DCNN is a Deep Neural Network with additional convolutional layers. It has an architecture lighter than other deep learning models. A large network having multiple convolutional layers presents this architecture. Figure 2 displays the proposed network that contains 7 layers. The first layer represents the input data. The second layer is a convolutional layer of size 256*256*3. Both Layer 3 and 5 are pooling layers using the function Softmax pooling w ith filter 3*3. The fourth layer is a convolutional layer with size 64*64*3. Layer 6 represents a fully connection layer having 4096 neurones. The last Layer (layer 7) is the output layer containing 1024 neurones where each neuron presents a low-level feature. For feature extraction, the values of last fully connected layer in DCNN model, represented but the output layer in Figure 2, is excluded from the network and combined into one vector of features. The dimension of this vector is 1024 that will be used to search the similar images. The feature extraction process is done offline to extract features from images and online to extract features from the query.

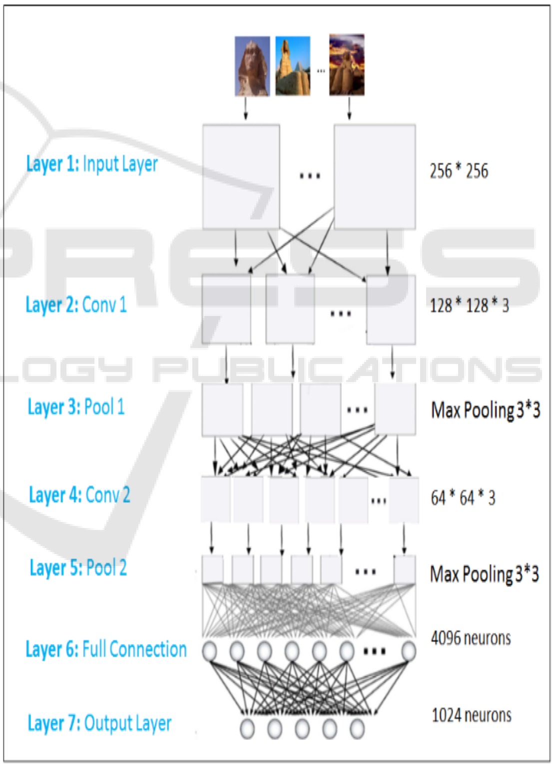

Figure 2: DCNN architecture.

2.2 Fuzzy C-Means Clustering Algorithm

After feature extraction, images from the dataset are clustered using a shallow learning technique “fuzzy c-means” that is given by Figure 3.

Fuzzy c-means is one of the most popular algorithm used in different area of research including computer science. It is the fuzzy version of k-means algorithm. FCM is used to make data clustering by a set of features and a number of initialized clusters. Let l is number of clusters randomly initialized. The target of FCM is to cluster the similar genes in a single cluster. Gene seems the vector of features. Let $X = \{x_i : i = 1, \dots, n\}$, where x_i is the value of gene g_i. The goal is to cluster all the gene into c clusters where $c \in \{2 \dots n-1\}$. The cluster c_j is assigned by a partition matrix $W = w_{g_i,j}$ that contain the degree of membership of gene g_i in cluster c_j.

For a cluster c_j, the corresponding centroid $Cent_j$ is defined as:

$$Cent_j = \frac{\sum_{i=1}^{n} w_{ij}^{p} x_i}{\sum_{i=1}^{n} w_{ij}^{p}} \quad (1)$$

Where $p \in [1..\infty[$ is a parameter that determines the influence of the weights $w_{g_i,j}$.

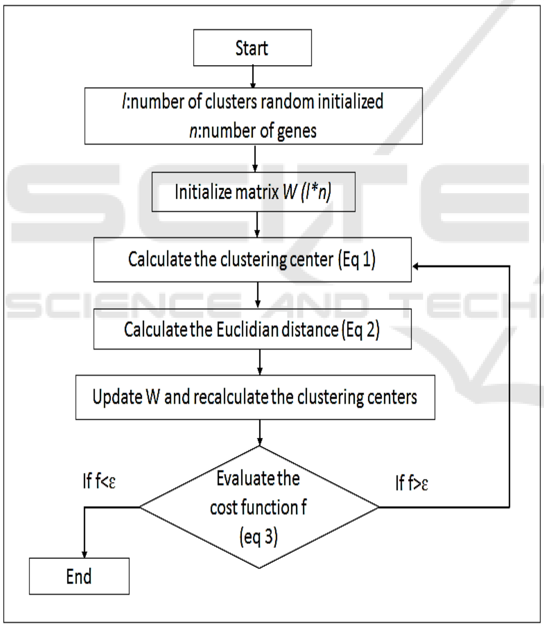

Figure 3: FCM algorithm.

The distance between a centroid ($Cent_j$) and a gene g_i represented by the features vector x_i, is calculated using Euclidean distance, as the following equation where z is the size of g_i:

$$dist(x_i, Cent_j) = \sqrt{\sum_{k=1}^{z} (x_{i_k}, Cent_{j_k})^2} \quad (2)$$

FCM attempts to minimize the cost function designer by the sum of the squared error (SSE), as with k-means algorithm.

$$f = \sum_{j=1}^{k} \sum_{i=1}^{n} w_{i,j}^{p} \, dist(x_i, Cent_j)^2 \quad (3)$$

After clustering, a label signature is assigned for each cluster. We means by label signature the signature of the centroid that is represented by vector of features.

2.3 Similarity Calculation

To search for images based on their low-level features, Euclidian distance (see equation 4, where k is the number of features and $j \in [0, c]$) is used. Similarity is calculated between the features of query and the label signatures of images. This phase is done online and all the obtained similarity scores are sorted in a descending order. The relevant label signature should have the first rank. Then the relevant cluster corresponding to its label is returned, and all the images belonging to that cluster are considered relevant.

$$Sim(q, Cent_j) = \sqrt{\sum_{k=1}^{z} (q, Cent_{j_k})^2} \quad (4)$$

On the other hand, images in the relevant cluster can be displayed in ascending order. In this case, a similarity score should be calculated between the label signature and the cluster images.

3 EXPERIMENTS

In this section, our contribution is evaluated using a set of experiments: the first set to evaluate the parameters of DCNN and the second set is proposed to compare the developed method with other state-of-the-art methods. The experiments are performed using two datasets:

INRIA Holidays dataset contains personal holiday's photos taken in different scene types. Images of each scene regroups almost 500 images. The first image (image number one) from each scene represents the query image.

Oxford5K dataset contains 5062 images selected from Flickr displayed a specific Oxford landmark. 11 different landmarks are regrouped as Ground truth where each one is represented by five possible queries. Therefore, number of queries equals 55.

Performance is evaluated using the mean Average Precision (mAP) function.

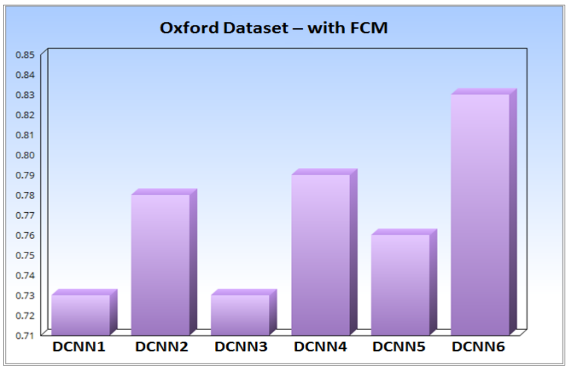

Figure 4: Results on Oxford with FCM.

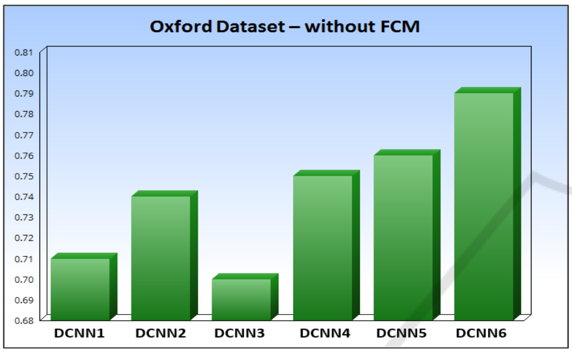

Figure 5: Results on Oxford without FCM.

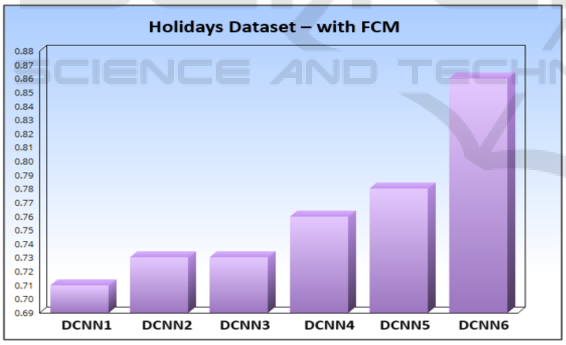

Figure 6: Results on Holidays with FCM.

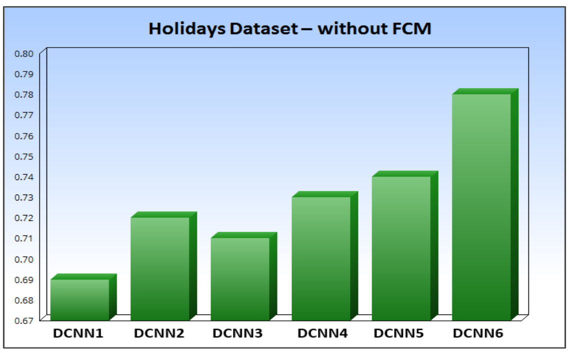

Figure 7: Results on Holidays without FCM.

3.1 Evaluation of DCNN for Feature Extraction and FCM for Clustering

In this section, we evaluate the proposed network, especially the DCNN and the FCM models. Concerning DCNN, we test six version of the proposed model. Each version includes different parameters including convolutional layer, pooling layers, fully connected layers and the activation function. In addition, each version contains 7 layers: input layer, 2 convolutional layers, 2 pooling layers, 1 layer full connection, and output layer. The six versions are:

DCNN1 stars with a convolutional layer having the size 224*224*3 and finishes with size equals to 14*14*512. The pooling layers include an average pooling function.
DCNN2 has identical set of parameters as DCNN1, but the pooling layers uses a Softmax pooling function.
DCNN3 begins with a convolutional layer of size 260*260*3 and output size equals to 48*48*512. The pooling layers include an average pooling function.
DCNN4 has the same parameters that DCNN3, but the pooling layers uses a Softmax pooling function.
DCNN5 is the proposed DCNN model described in the previous section but uses an average pooling function instead of Softmax pooling.
DCNN6 is the proposed DCNN model described in section 2.1.

We evaluate each version separately and we re-evaluate them with the proposed FCM model. The evaluation is done on two steps. First step concerns the search of queries using the features extracted from each model in an isolation mode. In second step evaluated the combination of that model with FCM.

Figures 4 and 5 display respectively the results on Oxford dataset with and without FCM and figures 6 and 7 represent respectively the results on Holidays dataset with and without FCM. The obtained results show that DCNN6 is the best model compared with other models containing or no FCM model. For the rest of paper, we complete experiments using DCNN6 model.

Table 1 represents the obtained results using Oxford and Holidays datasets using DCNN combined with or without FCM. On Oxford5k, the mAP performed 0.83 with the use of FCM against 0.79 without FCM. However, the values of mAP on Holidays are respectively 0.86 for DCNN and FCM and 0.78 using DCNN only. Therefore, we conclude that the use of FCM enhances the proposed model.

Table 1: Impact of FCM on Oxford and Holidays dataset when it is combined with DCNN6.

Method	Oxford5K	Holidays
DCNN+FCM	0.83	0.86
DCNN	0.79	0.78

3.2 Similarity Measure Evaluation

The choice of the similarity measure is an important step in a CBIR system. Therefore, to calculate the similarity between the label signatures and query we evaluated three distance measures: Chebyshev (equation 5), Manhattan (equation 6) and Euclidian (equation 4).

$$\mathrm{Chebyshev}(q, \mathrm{Cent}_j) = {}_{i=1}^{z}\max\left|q - \mathrm{Cent}_{j_k}\right| \quad (5)$$

$$\mathrm{Manhattan}(q, \mathrm{Cent}_j) = \sum_{k=1}^{z}\left|q - \mathrm{Cent}_{j_k}\right| \quad (6)$$

Table 2 concludes the obtained results on Oxford and Holidays using the three measures. From this table, we conclude that Euclidian distance gives the best results against the other distance measures. Is not the first time to show the performance of Euclidian distance in image retrieval. This measure is easy to use and very fast compared to other measures that need more computation time

Table 2: Evaluation of three similarity measures.

Measure	Oxford5K	Holidays
Chebyshev	0.79	0.8
Manhattan	0.77	0.82
Euclidian	0.83	0.86

3.3 (DCNN+FCM) versus Literatures

In this section, we compare the proposed model (DCNN and FCM) with other methods from literature that are recently developed. Figure 8, displays the obtained results for testing different deep learning models on Oxford5k and Holidays collections.

Using a basic Convolutional Neural Network (CNN) (Razavian et al., 2014), we obtained on Oxford mAP=32.2 and on Holidays, we obtained mAP=64.2. Changing the pooling function from average to max pooling (Razavian et al., 2016), the MAP can reach respectively 53.3 on Oxford and 71.6 on Holidays. NetVLAD (Arandjelovic et al.,

2016) is the system that combines CNN and VLAD, the mAP achieves 71.6 on Oxford5k and 87.5 on Holidays. The local descriptors like the bag of Word (Mohedano et al., 2016) can be merged to CNN to enhance the retrieval performance. Therefore, the mAP achieves 73.9 on Oxford. The model proposed by (Radenovic et al., 2017) (CNN-Gem) achieves the best results compared to the above methods and our method. The results are 87.8 on Oxford and 93.9 on Holidays.

Concerning our method, the obtained results are better than (Razavian et al., 2014, Razavian et al., 2016, Mohedano et al., 2016). This shows that the architecture of DCNN is well-suited feature extraction compared to CNN. In addition, we compared our results with (Arandjelovic et al., 2016) and (Radenovic et al., 2017) that have result more than we on the two datasets have. This difference is around 3% on Oxford and between 1% and 6% on Holidays. As recorded the difference between the proposed method and literatures is relatively small, however the proposed method over beats in the number of used parameters. It means, similar results are obtained with a lower cost and minimized

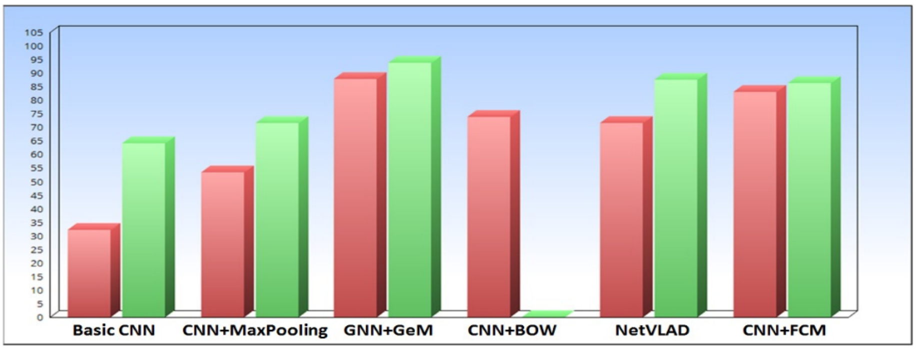

Figure 8: Comparison between our method and other methods from literature.

searching time (Arandjelovic et al., 2016, Radenovic et al., 2017) despite the fact that they have important values of mAP. In addition, it is not a benefit for such methods to be costly when we talk about real-time retrieving.

4 CONCLUSION AND FUTURE WORK

In this work, a new method for image retrieval based on the visual content of images was proposed. This method uses the DCNN technique for feature extraction. Then, it clusters the dataset and gives a label signature for each cluster. Finally, the similarity is calculated between the query and the labels to accelerate the retrieving process.

Performance was evaluated using two datasets Oxford5k and Holidays. The obtained results displayed the efficiency of the proposed method. Especially, when it was compared with other CBIR systems on from literature.

In future work, retrieving performance can be improved by the use of recent deep learning techniques like the Generative adversarial network (GAN).

REFERENCES

Arandjelovic, R., Gronat, P., Torii, A., Pajdla, T., Sivic, J., 2016. NetVLAD: CNN architecture for weakly supervised place recognition, *in CVPR*.

Babenko, A., Lempitsky, V., 2015. Aggregating local deep features for image retrieval, *in The IEEE International Conference on Computer Vision (ICCV).*

Babenko, A., Slesarev, A., Chigorin, A., Lempitsky, V., 2014. Neural codes for image retrieval. *In: European conference on computer vision*, Springer pp.584–599.

Fu, R., Li, B., Gao, Y., Wang, P., 2016. Content-based image retrieval based on CNN and SVM. *In: 2nd IEEE Int. Conf. on Computer and Communications (ICCC)*, pp.638–642.

Ghrabat, M.J.J., et al., 2019. An effective image retrieval based on optimized genetic algorithm utilized a novel SVM-based convolutional neural network classifier. Hum. *Cent. Comput. Inf. Sci.* 9, 31.

Gong, G., Wang, L., Guo, R., Lazebnik, S., 2014. Multi-scale orderless pooling of deep convolutional activation features. *Proceedings of the European Conference on Computer Vision*, pp.392-407.

Gopal, N., SBhooshan, R., 2015. Content based image retrieval using enhanced surf. In Fifth National Conference on Computer Vision, *Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp.1-4.

Gordo, A., Almazan, A., Revaud, J., Larlus, D., 2016. Deep image retrieval: Learning global representations for image search. *In European Conference on Computer Vision*, pp.241-257.

He, K., Zhang, X., Ren, S., Sun, J., 2014. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine*. vol.37, pp.1904-1916.

Hiremath, P.S., Pujari, J., 2007. Content based image retrieval using color, texture and shape features. *International conference on advanced computing and communications.* Guwahati, Assam, pp.780-784.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam. H., 2017. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications.

Jégou, H., Douze, M., Schmid, C., Pérez, P., 2010. Aggregating local descriptors into a compact image representation. *In Computer Vision and Pattern Recognition (CVPR)*, pp.3304-3311.

Karamti, H., 2013. Vectorisation du modèle d'appariement pour la recherche d'images par le contenu. *Conference en Recherche d'Infomations et Applications* (CORIA), pp.335-340.

Karamti, H., et al., 2018. Vector space model adaptation and pseudo relevance feedback for content-based image retrieval. *Multimedia Tools and Applications*. vol.77, pp.5475-5501.

Kim, W., Goyal, B., Chawla, K., Lee, J., Kwon, K., 2018. Attention-based ensemble for deep metric learning. *In The European Conf on Computer Vision (ECCV).*

Lin, Z., Yang, Z., Huang, F., Chen, J., 2018. Regional maximum activations of convolutions with attention for cross-domain beauty and personal care product retrieval. *ACM Multimedia Conf*., pp.2073–2077.

Mohedano, E., Salvador, A., McGuinness, K., Marques, F., O'Connor, NE., A. Salvador, F., Giro-i Nieto, G., 2016. Bags of local convolutional Features for Scalable Instance Search.

Nazir, A., Ashraf, A., Hamdani, T., Ali, N., 2018. Content based image retrieval system by using hsv color histogram, discrete wavelet transform and edge histogram descriptor. *In International Conference on Computing. Mathematics and Engineering Technologies (iCoMET)*, pp.1-6.

Nguyen, N., Rigaud. C., Burie. JC., 2018. Digital Comics Image Indexing Based on Deep Learning. *J. Imaging*. vol.4(7), pp.89.

Paulin, M., et al., 2015. Local convolutional features with unsupervised training for image retrieval. *In ICCV*.

Piras, L., Giacinto, G., 2017. Information fusion in content based image retrieval: A comprehensive overview. *Information Fusion*. vol.37, pp.50-60.

Qassim, H., Verma, A., Feinzimer, D., 2018. Compressed Residual-VGG16 CNN Model for Big Data Places Image Recognition. *IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*.

Qin, J.H., et al.,., 2015. Ceramic tile image retrieval method based on visual feature. *Journal of*

Computational and Theoretical Nanoscience. vol.12, pp.4191-4195.

Radenovic, F., Tolias, G., Chum, O., 2017. Fine-tuning CNN Image Retrieval with No Human Annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S., 2014. CNN features off-the-shelf: an astounding baseline for recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp.806-813.

Razavian, A.S., Sullivan, J., Carlsson, S., Maki, A., 2016. Visual instance retrieval with deep convolutional networks. *ITE Transactions on Media Technology and Applications*, vol. 4(3), pp.251-258.

Sun, G., Wang, C., Ma, B., Wang, X., 2017. An improved SIFT algorithm for infringement retrieval. *Multimedia tools and Applications*.

Varish, K.N., Pal, A.K., 2015. Content based image retrieval using statistical features of color histogram. *In International Conference on Signal Processing*, Communication and Networking (ICSCN), pp.1-6.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.A., 2010. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res*. vol. 11, pp.3371-3408.

Wang, J.Z., Li, J., Wiederhold, G., 2001. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol.23 (9), pp.947-963.

Wu, L., Wang, Y., Gao, J., Li, X., 2018. Deep adaptive feature embedding with local sample distributions for person re-identification. *Pattern Recognition*. vol.73, pp.275-288.

Wu, L., Wang, Y., Li, X., Gao, J., 2018. Deep Attention-Based Spatially Recursive Networks for Fine-Grained Visual Recognition. *IEEE Transactions on Cybernetics*. vol.49, pp.1791-1802.

Wu, Y., Wu, Y., 2009. Shape-Based Image Retrieval Using Combining Global and Local Shape Features *International Congress on Image and Signal Processing*, Tianjin, pp.1-5.

Yuan, X., Yu, J., Qin, Z., Wan, T., 2011. A sift-lbp image retrieval model based on bag-of- features. *Int. conference on image processing*, pp.1061-1164.

Yu, J., Qin, Z., Wan, T., Zhang, X., 2013. Feature integration analysis of bag-of-features model for image retrieval. *Eurocomputing*, vol. 120, pp. 355-364.

Zhou, W., Li, H., Tian, Q., 2019. Recent Advance in Content-based Image Retrieval: A Literature Survey.