

## Trademark Image Retrieval using Transfer Learning

Shahla J. Hassen, Ahmed Taha and Mazen M. Selim  
Department of Computer Science, Faculty of Computers and Informatics,  
Benha University, Benha, Egypt, eng.shahla@yahoo.com

**Abstract:** Trademarks are valuable assets that need to be protected from infringement for the sake of producers and consumers. Therefore, Trademark Image Retrieval (TIR) is getting an increasing attention both academically and commercially. Recently, convolutional neural networks have stand out as a compulsory alternate. It offers perfect predictive performance and the possibility to replace classical workflows with an only network architecture. In addition, the transfer learning can save time and efforts in building deep convolutional neural networks. In this study, a transfer learning based TIR system is presented. It employs AlexNet, a pre-trained deep convolutional neural network. The proposed system is evaluated and validated using the two benchmark datasets: “FlickrLogos32” and “Logos-32 Plus” in terms of well-known performance metrics. The obtained results show that our proposed system has a promising performance compared to other recent systems.

**Key words:** Content-based image retrieval, trademark image retrieval, deep learning, transfer learning, AlexNet, academically

### INTRODUCTION

Content Based Image Retrieval (CBIR) is a signal and multimedia processing research field with many promising applications. CBIR is a dynamic area of research with few years of attention. Recently, there exist many research subfields that are emerged from the CBIR field. This study concentrates on one a subset of CBIR, namely logo and trademark retrieval. It is receiving an increasing attention in the literature due to its relevance to many promising, commercial, real-world applications. It is considered an important tool for protecting the rights of the companies that work in industry, commerce and sports entertainment (Phan and Androutsos, 2010).

A trademark is a mark that identifies one person's goods or services. In practice, the word “trademark” is used to refer to any class of mark or a word, symbol or phrase used to distinguish a particular manufacturer's product from those of others (Her *et al.*, 2011). As shown in Fig. 1, there are four types of trademark images including the word-in mark, the device mark, the composite-mark and the complex mark. The word-in mark refers to the marks that contain only characters or word while the device mark refers to the marks that includes graphical shapes. The composite mark is a combination of the previously mentioned types. Finally, the complex mark is simply a composite mark with additional color and visual effects (Anuar *et al.*, 2013; Wei *et al.*, 2009).



Fig. 1: Examples for the different types of trademarks

Registered logos are considered a type of legitimate property that need protection from brand piracy and trademark infringement. In order to protect and make their trademarks legitimate, owners should record their trademark in patent offices in many countries worldwide. There exist around 100 million companies that work in local and global markets (Tursun *et al.*, 2017). Many of them have at least one registered trademarks. According to Word Intellectual Property Organization, there are three million trademark registrations all over the world. Additionally, trademark registration requests are increasing at a rate 6-8% in recent years (Tursun *et al.*, 2017).

During the registration of a new trademark, it is necessary to be sure that the new trademark does not mimic or is similar enough to existing trademarks. When the trademarks are not comparable, similarity between them is measured using the likelihood of confusion combined with the belief that a product or service is offered by the company that owns the registered

trademark. Lawfully, the likelihood of confusion is assessed using a number of factors. One of these factors is the visual similarity of the trademarks (Anuar *et al.*, 2013). One of the steps performed during the registration process is to make sure that the trademark that need to be registered is not similar to any previously registered trademark. This is important to protect the rights of the owners of the already registered trademark in addition to prevent infringements (Anuar *et al.*, 2013). In general, registering the trademark images is beneficial for both traders and clients. Traders distinguish and protect their services or products by a unique mark while clients can be assured about the originality of the goods by its trademark (Kochakornjarupong, 2001).

In most developed countries, organizations like patent offices are responsible for preventing trademarks infringements. They refuse to register the new trademarks that are similar enough to already registered ones either checking the trademarks database manually or by employing a trademark retrieval system. Toward achieving this mission, many approaches have been developed starting from manual approaches ending with fully automated retrieval systems (Laiwen *et al.*, 2014). However, the huge number of registrations have badly affected the performance of both manual and automatic operations. It decreased the quality of the service provided by patent offices which left a big room for trademark infringements. Moreover, mistakenly registering two similar trademarks will complicate the handling of legal disputation between the owners of the trademarks. In order to reduce the burdens of patent offices, an efficient and accurate automated Trademark Retrieval (TR) system with intelligent image analysis techniques is a necessity (Tursun *et al.*, 2017). In this study an automated TIR system based on the transfer learning approach is presented. It uses AlexNet, the well-known Deep Convolutional Neural Network (DCNN). The proposed system comprises three main steps: image preprocessing, network tuning and network training.

**Literature review:** Generally, many methods have been presented in the field of trademark image retrieval. Researchers have been working on developing fully automated methods for similar trademark retrieval for around two decades. However, this field still needs many efforts. The traditional trademark images retrieval systems may involve a number of steps such as logos segmentation, feature extraction, feature reduction, feature optimization and classification, similarity calculation (Tursun *et al.*, 2017; Laiwen *et al.*, 2014; Aires *et al.*, 2015; Tursun and Kalkan, 2015; Anuar *et al.*, 2016; Yan *et al.*,

2016). On the other side, recently, the concept of end-to-end learning has been employed in many fields including trademark image retrieval (Oliveira *et al.*, 2016; Pinjarkar *et al.*, 2018). In the latest approach, the deep neural networks work directly on the pixel's values of trademark images without the need to perform any additional tasks such as logo segmentation or feature extraction. In this study, some of the efforts that have been done in the field are reviewed briefly.

Laiwen *et al.* (2014) present a TIR system that employs both text-based retrieval and content-based image retrieval technologies. The text-based retrieval module performs a number of functionalities including fuzzy retrieval, precise retrieval, etc. In addition, the content-based image retrieval module can extract variety of features including color, texture and/or shape features from trademark images. The similarity calculation is based on the Euclidean distance. The obtained results on a proprietary dataset have shown that the system has 77% precision rate and 95.1% recall rate in the best case (Laiwen *et al.*, 2014).

Another TIR system is presented by Aires *et al.* (2015) in which the trademark images are distinguished using Scale Invariant Feature Transform (SIFT). Moreover, the system represents the shape's visual perception aspects using non-symmetrical perceptual zoning mechanism that depends on the Principles of Gestalt. The system has been assessed using a trademark images dataset called UK database patent office. It contains 10745 logos in grayscale. The obtained results have shown that the achieved normalized recall is 84% (Aires *et al.*, 2015).

Furthermore, Tursun and Kalkan (2015) build a trademark image dataset (METU dataset) that includes more than 930,000 trademarks. In addition, they evaluate many of the current logo retrieval approaches using their dataset. The conducted experiments show that the other approaches are not good enough with challenging datasets such as the suggested one (Tursun and Kalkan, 2015). Moreover, Bianco *et al.* (2017) address some issues in trademark image retrieval including the reversal of contrast and presence of irrelevant text in trademarks. They employ the deep learning for the first time in trademark image retrieval.

Another relevant work that addresses the trademark image retrieval problem from a different perspective is presented by Anuar *et al.* (2016). The researcher measure the conceptual similarity between trademarks based on the semantic content. Their system employs NLP approaches and external knowledge source represented by a lexical ontology. Also, it measures the similarity distance based on Tversky's theory of similarity. The

performance of the system is assessed in terms of R-precision score and human judgment. It uses a trademark database of 1400 disputed cases in addition to another database of 378943 company names. The obtained results have shown that the system has 82% R-precision in the best case.

Additionally, Yan *et al.* (2016) propose an adaptive fusion approach for color and spatial descriptors for colored trademark images retrieval. First, the effective dominant color is extracted, then, the spatial descriptor is built for the extracted dominant color. Next, an adaptive fusion is applied on these two features. The system has been evaluated using colored trademark images dataset of more than 2300 trademarks. The achieved results have shown the superiority of their approach over a number of other recent trademark retrieval approaches (Yan *et al.*, 2016).

Oliveira *et al.* (2016) present a logo recognition and retrieval system that depends on Fast Region-based Convolutional Networks (FRCN) approach. The researcher employ two CNN models that have pre-trained using ILSVRC ImageNet dataset. In addition, they look at the selective search of windows 'proposal's during the pre-processing and data augmentation to increase, the recognition rate. The system is evaluated using FlickrLogos-32 dataset and the experimental results show that the system has 95.5% precision, 90.8% recall and 93.1% F1-measure (Oliveira *et al.*, 2016).

Pinjarkar *et al.* (2018) propose another logo recognition and retrieval system that adopts both optimization and unsupervised learning during the pre-processing phase. The Particle Swarm Optimization (PSO) is employed to reduce the search space through optimizing the database feature set. Then, the optimized feature set is clustered using the Self-Organizing Map (SOM) approach. In addition in order to perform the recognition and retrieval tasks with high performance a Relevance Feedback (RF) is embedded with both short-term learning and long-term learning. The system is evaluated with a proprietary dataset called "FlickrLogos-32 Plus". The obtained results have shown that the proposed system has 96.8% mean average precision (map) and 93.12% accuracy (Pinjarkar *et al.*, 2018).

Further researcher is provided by Liu *et al.* (2017) in the field of logo recognition and retrieval. The system employs the Scale Invariant Feature Transform (SIFT) detectors as the basic feature extraction method. Then, it uses the generalized Hough transform to add spatial information into sets of local features through finding local features that are invariant across images. In addition, it adopts a reference-based image representation

approach. Finally, a Support Vector Machine (SVM) classifier with various kernel functions. The proposed system is evaluated using FlickrLogos-32 dataset. The obtained results show that the best performance is achieved using the linear kernel function with 96.2% precision and 86.4% recall (Liu *et al.*, 2017).

Additionally, Bianco *et al.* (2017) present a deep learning-based a logo recognition system. The proposed system employs a logo region proposal approach that is highly recall-oriented. Then, a Convolutional Neural Network (CNN) is used to perform the classification step. Moreover, the benefits of applying a number of machine learning techniques such as pre-processing, class-balancing, sample weighting are investigated. The proposed system is trained using FlickrLogos-32 dataset and Logos-32 Plus dataset. On the other side, the performance of the system is evaluated only using the FlickrLogos-32 dataset. The obtained results show that the proposed system has 98.9% precision, 91.7% recall and 96% accuracy (Bianco *et al.*, 2017).

**Transfer learning:** Logos recognition and retrieval problem is a special case of the general problem of object recognition (Iandola *et al.*, 2015). Although, conventional approaches for logo classification and retrievals have shown positive results (Tursun *et al.*, 2017; Laiwen *et al.*, 2014; Aires *et al.*, 2015; Tursun and Kalkan, 2015; Anuar *et al.*, 2016; Yan *et al.*, 2016), they involve many non-trivial data analysis steps. These steps may include ROI segmentation, feature extraction, feature selection or optimization and classification.

Recently, Deep Convolutional Neural Networks (DCNNs) have made breakthroughs in the fields of computer vision and image processing. DCNNs can classify images using the automatically extracted features that only depends on the raw pixel's intensity values. This supervised learning approach for extracting the features automatically have shown its superiority compared to classical hand-crafted features. Merging the segmentation and classification processes in a single framework made the image classification possible without the need to a separate ROI segmentation step. That eliminated a step that often needs a special caring and a large amount of computation (Kensert *et al.*, 2018). DCNNs have been employed successfully in many classification problems. For example, in the ImageNet object classification challenge, DCNNs have improved the classification accuracy over years with good rates. Also, by employing DCNNs, Donahue *et al.* (2014) have achieved better classification accuracy on scene classification and fine-grained bird classification

compared to the state-of-the-art approaches. In addition, DCNNs enabled Sharif Razavian *et al.* (2014) achieve better accuracy in human attribute detection and visual instance retrieval compared to the state-of-the-art approaches. Rather than building a deep network from scratch, the proposed system adopts the principal of transfer learning through using a deep network that already achieved success in some field, then, tuning this network to perform the trademark images recognition. During surfing the literature for DCNNs architectures, three major DCNNs have been found, namely AlexNet (Krizhevsky *et al.*, 2012), VGG-19 (Simonyan and Zisserman, 2014) and GoogLeNet (Szegedy *et al.*, 2015).

AlexNet, Donahue *et al.* (2014) is a DCNN that is designed and presented by Krizhevsky *et al.* (2012) for the ImageNet-2012 image classification challenge. It won this challenge achieving great improvements in the classification accuracy compared to the previous years. Its architecture includes eight main layers namely, five convolutional layers and three fully-connected layers. By 2014, state-of-the-art ImageNet DCNNs had evolved to be more deeper networks that contains up to nineteen layers such as VGG-19 (Simonyan and Zisserman, 2014). In addition to increasing depth, GoogLeNet (Szegedy *et al.*, 2015) had added what called “inception” meta-layers that

include multiple convolution filter resolutions. The objective of the addition of the inception layers to make the network able to accurately classify the objects regardless its size with respect to the overall image size. GoogLeNet and VGG-19 took the top two places in the ImageNet-2014 competition. However, due to the huge computing power needed for using GoogLeNet and VGG-19, AlexNet is employed in the proposed system for performing trademark image retrieval and recognition.

## MATERIALS AND METHODS

**Proposed system:** In this study, the proposed trademark image retrieval system is presented. It three main phases: dataset preprocessing, network tuning and network training. The block diagram of the proposed system is shown in Fig. 2. The details of each step are given below in a dedicated subsection.

**Preprocessing:** In the preprocessing step, all the images are resized to be  $227 \times 227$  pixel. In addition, all the trademark images are assured to be in RGB format and the images of other formats are transformed into RGB format. Figure 3 shows a trademark image before and after applying the preprocessing step.

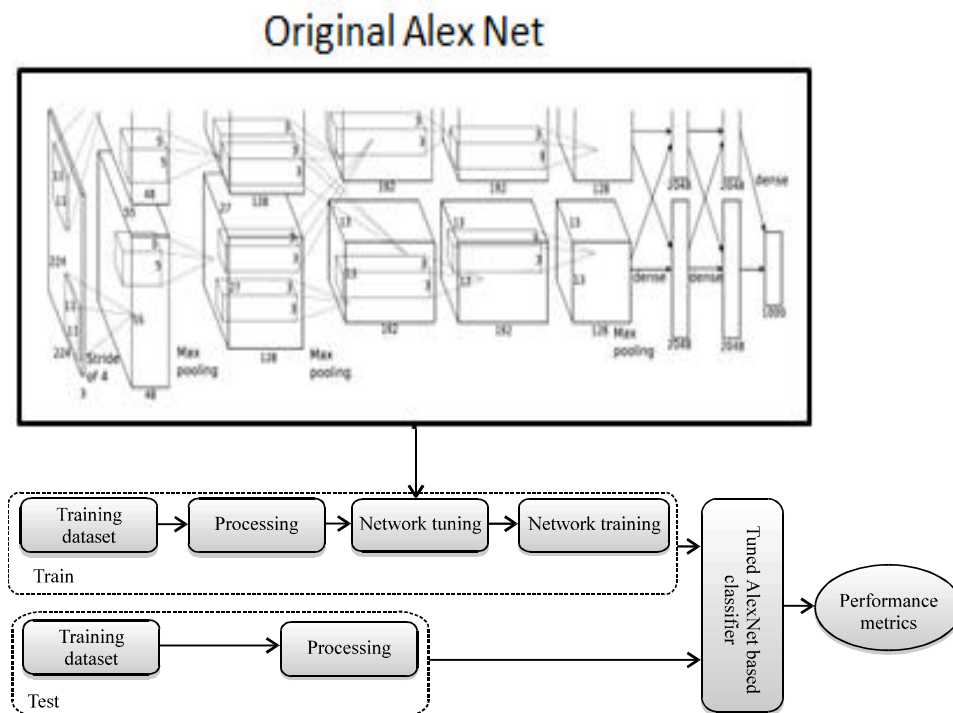


Fig. 2: The block diagram of the proposed trademark image retrieval system

(a)



(b)



Fig. 3: A trademark image before and after the preprocessing step: a) Original image and b) Preprocessed image

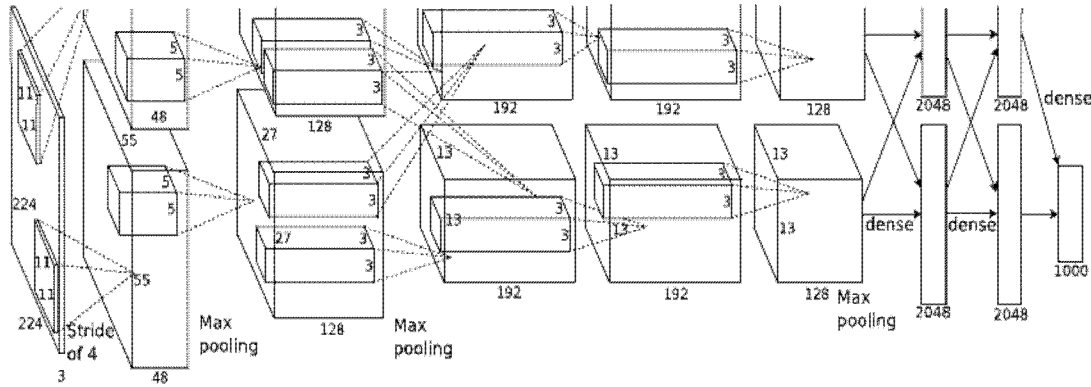


Fig. 4: The original architecture for AlexNet network (Krizhevsky *et al.*, 2012)

**Network tuning:** The original architecture of AlexNet that was proposed for the ImageNet object classification challenge is shown in Fig. 4. It comprises five convolution layers and three fully connected layers. ILSVRC challenge has used a subset of the ImageNet dataset that contains 1000 class and 1000 image. AlexNet has employed the Rectified Linear Unit (ReLU) function as an activation function. It is characterized by its high speed compared to other activation functions such as hyperbolic tangent function or sigmoid function. The equation of ReLU function is given below (Krizhevsky *et al.*, 2012):

$$f(x) = \max(0, x) \quad (1)$$

In order to increase the generalization ability of the net, local response normalization is applied where the response-normalized activity  $b_{x,y}^i$  can be computed using the following Eq. 2 (Krizhevsky *et al.*, 2012):

$$b_{x,y}^i = \frac{a_{x,y}^i}{k + \alpha \sum_{j=\max(0, i-\frac{n}{2})}^{\min(N-1, i+\frac{n}{2})} (a_{x,y}^j)^2}^\beta \quad (2)$$

where,  $a_{x,y}^i$  refers to the activity of a neuron computed by applying kernel  $i$  at the location  $(x, y)$  after performing the ReLU nonlinearity,  $n$  refers to the number of the adjacent kernel maps on which the sum runs,  $N$  refers to the total number kernel maps exist in the layer and  $k, n, \alpha$  and  $\beta$  are hyper-parameters (Krizhevsky *et al.*, 2012).

Also, pooling layers are employed to summarize the outputs of neighboring neurons in the same kernel map. Finally, the over-fitting problem is handled by adopting two approaches namely, data augmentation and dropout. In data augmentation approaches, the over-fitting problem can be handled through two forms. In the first form, image translations and horizontal reflections are generated. In the second form, the intensities of the RGB channels are





Fig. 5: A visual summary for the FlickrLogos32 dataset

Dataset	No of brands	Total number of images
FlickrLogos-32	32	8240
Logos-32 plus	32	12312

altered. On the other side in the dropout approach the output of each hidden neuron is set to zero with probability 0.5. The dropout process is applied on the first two fully-connected layers. In the original AlexNet, the output of the last fully-connected to as input to a 1000-way softmax that generates a distribution over the 1000 class labels (Krizhevsky *et al.*, 2012).

In the network tuning step, AlexNet is tuned to cope with our requirements. The tuning process involves replacing the last three layers of the original AlexNet namely, the last fully-connected layer, the softmax layer and the output layer with another three layers that match with the proposed work. The last fully-connected layer is replaced because the weights contained in the layers are doing well with the Image Netchallenge not with the trademark images recognition and retrieval, therefore, the fully connected layer is replaced with a new one in which the weights will be computed though a training process. Also, the soft-max layer of the original AlexNet was designed with 1000-way where the ImageNet challenge include 1000 classes, therefore, it is replaced with a soft-max layer with 32-way that achieve the requirements of the proposed work where the used datasets include 32 classes. Similarly, the output layer of the original AlexNet was designed to be able to generate 1000 different outputs that corresponds to the 1000 classes of the ImageNet challenge; therefore, the output layer is redesigned to be able to generate 32 different outputs that

correspond to the 32 classes used in the proposed work. The rest of the AlexNet architecture is kept untouched with the same weights gained during training the network for the ImageNet challenge.

**Network training:** In this step, the tuned network is trained using the logos datasets. In the conducted training process, the batch size is 10, the number of epochs per iteration is 6, the learning rate is 0.2, the value of bias is 0.2 and the data augmentation's translations and reflections are performed within.

**System implementation and evaluation:** In this study, the proposed system is evaluated using the FlickrLogos32 (Anonymous, 2018) and Logos-32 Plus (Bianco *et al.*, 2017) datasets. First, the used dataset is described in detail. Then, the experiments are presented and the obtained results are analyzed.

**Dataset description:** To evaluate the proposed system, we have employed two publically available datasets namely "FlickrLogos32" dataset (Anonymous, 2018) and "Logos-32 Plus" dataset (Bianco *et al.*, 2017). FlickrLogos32 data consists of images showing a number of brand logos, such as Adidas, Aldi, Apple, etc., that can be used for evaluating logo retrieval systems. It includes 32 logo classes, each have 70 images. In addition, FlickrLogos32 includes a class called Non-logo class that contains 6000 images yielding a dataset of 8240 total number of images. A visual summary of the used dataset is shown in Fig. 5. On the other hand, "Logos-32 Plus" dataset is an extension of the training-set of "FlickrLogos

```
>> net.Layers

ans =

25x1 Layer array with layers:

    1  'data'      Image Input      227x227x3 images with 'zerocenter' normalization
    2  'conv1'     Convolution      96 11x11x3 convolutions with stride [4 4] and padding [0 0 0 0]
    3  'relu1'     ReLU              ReLU
    4  'norm1'     Cross Channel Normalization cross channel normalization with 5 channels per element
    5  'pool1'     Max Pooling        3x3 max pooling with stride [2 2] and padding [0 0 0 0]
    6  'conv2'     Convolution      256 5x5x48 convolutions with stride [1 1] and padding [2 2 2 2]
    7  'relu2'     ReLU              ReLU
    8  'norm2'     Cross Channel Normalization cross channel normalization with 5 channels per element
    9  'pool2'     Max Pooling        3x3 max pooling with stride [2 2] and padding [0 0 0 0]
   10  'conv3'     Convolution      384 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
   11  'relu3'     ReLU              ReLU
   12  'conv4'     Convolution      384 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
   13  'relu4'     ReLU              ReLU
   14  'conv5'     Convolution      256 3x3x192 convolutions with stride [1 1] and padding [1 1 1 1]
   15  'relu5'     ReLU              ReLU
   16  'pool5'     Max Pooling        3x3 max pooling with stride [2 2] and padding [0 0 0 0]
   17  'fc6'       Fully Connected    4096 fully connected layer
   18  'relu6'     ReLU              ReLU
   19  'drop6'     Dropout            50% dropout
   20  'fc7'       Fully Connected    4096 fully connected layer
   21  'relu7'     ReLU              ReLU
   22  'drop7'     Dropout            50% dropout
   23  'fc'        Fully Connected    4 fully connected layer
   24  'softmax'   Softmax              softmax
   25  'classoutput' Classification Output crossentropyx with 'EOSINOPHIL' and 3 other classes
```

Fig. 6: The details and order of the layers contained in the used net

32". It has the same number of logo classes but with a larger number of images contained in each class (12312 instances). The main objective during collecting the images of Logos-32 Plus dataset was to build a dataset that contains a larger set of images that were captured under real imaging conditions and have transformations that may happen in uncontrolled acquisition environment. The dataset comprises on average 400 images per class. Each image has a least one instance of the same class. A summary for the used logo datasets is given in Table 1.

**System implementation:** The proposed system has been implemented using MATLAB 2018a. The constructed network consists of 25 layers. The network comprises eight layers with weights namely, five convolutional layers and three fully-connected layers. The output of the fully connected layer is as input to 32-way softmax layer. The first and second convolutional layers are followed by response normalization layers. The max-pooling layers employed in the used net follow the response normalization layers and the last convolutional layer. The ReLU is applied to the outputs generated from each convolutional and fully-connected layer. The dropout

layers follow the first two fully-connected layers. Hence, the constructed network contains one input layer, five convolutional layers, three fully-connected layers, two response normalization layers, three max-pooling layers, seven ReLU layers, two dropout layers, one softmax layer and one output layer. Each fully-connected layer has 4096 neurons, therefore, the obtained feature vectors are of 4096 length. The summary of the net layers and their order is shown in Fig. 6.

## RESULTS AND DISCUSSION

**Experiments and results analysis:** The proposed logos recognition and retrieval system has been evaluated in terms of a number of performance metrics on two logos datasets: FlickrLogos32 dataset (Anonymous, 2018) and Logos-32 plus dataset (Bianco *et al.*, 2017). The used performance metrics include the classification accuracy, recall and precision which are defined (Pinjarkar *et al.*, 2018):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Table 2: The performance of the proposed system using different test data ratios

Test data ratio (%)	FlickerLogos32			Logos-32 Plus		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
10	98.78641	98.9377	97.8355	87.86718	95.5389	91.8154
15	98.85246	98.9097	97.8754	88.23529	95.8119	92.0575
20	98.48301	98.7212	97.3642	87.79553	95.2085	91.9654
25	98.33659	98.6126	96.7540	87.87103	95.1276	92.2428
30	97.85599	98.0441	95.9096	87.73946	95.2684	92.2089
35	97.87308	97.9523	96.0357	88.21598	95.4655	92.5336
40	97.90655	97.9861	96.2603	88.02682	95.3957	92.6003
45	97.86024	97.7353	96.1351	87.62770	95.3539	92.2740
50	97.79126	97.6204	96.1396	87.49361	95.4489	92.0698
Max.	98.85246	98.9377	97.8754	88.23529	95.8119	92.6003
Min.	97.79126	97.6204	95.9096	87.49361	95.1276	91.8154
Avg.	98.19395	98.2799	96.701	87.87473	95.4022	92.1964

Table 3: Comparison of logo recognition methods with the proposed method

Method	Dataset	Performance metrics (%)		
		Accuracy	Precision	Recall
Oliveira <i>et al.</i> (2016)	FlickerLogos-32	-	95.50	90.80
Liu <i>et al.</i> (2017)	FlickerLogos-32	-	96.20	86.40
Bianco <i>et al.</i> (2017)	FlickerLogos-32	96.00	98.90	91.70
Proposed work	FlickerLogos-32	98.19	98.27	96.70
Proposed work	Logos-32 Plus	87.87	95.40	92.19

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (5)$$

Where:

TP = The True Positive

TN = The True Negative

FP = The False Positive

FN = The False Negative

The performance of the proposed system is evaluated under different scenarios of different train data/test data splitting ratios. The obtained results for the different performance metrics are shown in Table 2.

Based on Table 2 using “FlickerLogos32” dataset, the best achieved results of the proposed system regarding the classification accuracy and recall are achieved using 15% test data ratio with 98.85% accuracy and 97.87% recall while the best achieved results of the proposed system regarding the precision is achieved using 10% test data ratio with 98.93% precision. However, the average performance of the proposed system is 98.19, 98.27 and 96.7% for the accuracy, precision and recall, respectively.

On the other hand, using “Logos-32 Plus” dataset, the best achieved results of the proposed system regarding the classification accuracy and precision are achieved using 15% test data ratio with 88.23% accuracy and 95.81% precision while the best achieved results of the proposed system regarding the recall is achieved

using 40% test data ratio with 92.6% precision. However, the average performance of the proposed system is 87.87, 95.4 and 92.19% for the accuracy, precision and recall, respectively. Also, the proposed system has been compared to other recently proposed logo recognition and retrieval systems. The comparison is presented in Table 3.

Based on Table 3, when we compare the proposed work and the researcher present by Oliveira *et al.* (2016) and Liu *et al.* (2017) on the “FlickerLogos-32” dataset, it is noticed that the proposed method is better in terms of both precision and recall. Moreover, when we compare the proposed researcher and the work present by Bianco *et al.* (2017) it is noticed that the proposed researcher is better in terms of both accuracy and recall, while the researcher present by Bianco *et al.* (2017) is better in terms of precision.

## CONCLUSION

Logo recognition and retrieval is an important and active research field due to its relation to many real-world applications. Traditional logo recognition approaches have many challenges where logos may seem in any position, scale and under any point of view in an image. In this study, we have proposed a logo recognition and retrieval system that applies the transfer learning concept. The proposed system depends on a well-known deep convolutional neural network called AlexNet that achieved a big success in the ImageNet challenge. The proposed system comprises of three main steps: image preprocessing, network tuning and network training. It has been evaluated using two logos dataset called “FlickerLogos-32 dataset” and “Logos-32 Plus” dataset in terms of accuracy, precision and recall. The best results of the system has been obtained using “FlickerLogos-32” 98.19, 98.27 and 96.7% for the accuracy, precision and recall, respectively. Also, the proposed system has been compared to other recently proposed logo recognition and retrieval systems on the same dataset and the obtained results have revealed the superiority of the proposed system.

## REFERENCES

- Aires, S.B.K., C.O.A. de Freitas and L.S. Oliveira, 2015. SIFT applied to perceptual zoning for trademark retrieval. Proceedings of the 2015 IEEE International Conference on Systems, Man and Cybernetics, October 9-12, 2015, IEEE, Kowloon, China, ISBN:978-1-4799-8697-2, pp: 2401-2406.
- Anonymous, 2018. Datasets: Flickr Logos-32/Flickr Logos-47. University of Augsburg, Augsburg, Germany. <http://www.multimedia-computing.de/flickrlogos/#flic%20krlogos47>



- Anuar, F.M., R. Setchi and Y.K. Lai, 2013. Trademark image retrieval using an integrated shape descriptor. *Exp. Syst. Appl.*, 40: 105-121.
- Anuar, F.M., R. Setchi and Y.K. Lai, 2016. Semantic retrieval of trademarks based on conceptual similarity. *IEEE. Trans. Syst. Man Cybern. Syst.*, 46: 220-233.
- Bianco, S., M. Buzzelli, D. Mazzini and R. Schettini, 2017. Deep learning for logo recognition. *Neurocomputing*, 245: 23-30.
- Donahue, J., Y. Jia, O. Vinyals, J. Hoffman and N. Zhang *et al.*, 2014. Decaf: A deep convolutional activation feature for generic visual recognition. *Proceedings of the 31st International Conference on Machine Learning (PMLR)* Vol. 32, June 21-26, 2014, Beijing, China, pp: 647-655.
- Her, I., K. Mostafa and H.K. Hung, 2011. A hybrid trademark retrieval system using four-gray-level zernike moments and image compactness indices. *Intl. J. Image Proc.*, 4: 631-646.
- Iandola, F.N., A. Shen, P. Gao and K. Keutzer, 2015. Deeplogo: Hitting logo recognition with the deep neural network hammer. *Comput. Vision Pattern Recogn.*, 1: 1-12.
- Kensert, A., P.J. Harrison and O. Spjuth, 2018. Transfer learning with deep convolutional neural networks for classifying cellular morphological changes. *Slas Discovery Adv. Life Sci.*, 1: 1-10.
- Kochakornjarupong, P., 2001. Trademark image retrieval by local features. Ph.D Thesis, College of Social Sciences, University of Glasgow, Scotland, UK.
- Krizhevsky, A., I. Sutskever and G.E. Hinton, 2012. Image net classification with deep convolutional neural networks. *Proc. Neural Inf. Process. Syst.*, 1: 1097-1105.
- Laiwen, Y., S. Ping and L. Yanhong, 2014. Design and implementation of trademark image retrieval system. *Proceedings of the 2014 International Conference on Audio, Language and Image Processing*, July 7-9, 2014, IEEE, Shanghai, China, ISBN:978-1-4799-3902-2, pp: 160-165.
- Liu, Y., J. Wang, Z. Li and H. Li, 2017. Efficient logo recognition by local feature groups. *Multimedia Syst.*, 23: 395-403.
- Oliveira, G., X. Frazao, A. Pimentel and B. Ribeiro, 2016. Automatic graphic logo detection via fast region-based convolutional networks. *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, July 24-29, 2016, IEEE, Vancouver, British Columbia, Canada, ISBN:978-1-5090-0621-2, pp: 985-991.
- Phan, R. and D. Androutsos, 2010. Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient Co-occurrence histograms. *Comput. Vision Image Understanding*, 114: 66-84.
- Pinjarkar, L., M. Sharma and S. Selot, 2018. Novel relevance feedback approach for color trademark recognition using optimization and learning strategy. *J. Intell. Syst.*, 27: 67-79.
- Sharif Razavian, A., H. Azizpour, J. Sullivan and S. Carlsson, 2014. CNN features off-the-shelf: An astounding baseline for recognition. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPR)*, June 23-28, 2014, Columbus, Ohio, USA., pp: 806-813.
- Simonyan, K. and A. Zisserman, 2014. Very deep convolutional networks for large-scale image recognition. *J. Comput. Vision Pattern Recognit.*, 1: 1-14.
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet and S. Reed *et al.*, 2015. Going deeper with convolutions. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2015)* Vol. 2, June 7-12, 2015, Boston, Massachusetts, USA., pp: 1-9.
- Tursun, O. and S. Kalkan, 2015. METU dataset: A big dataset for benchmarking trademark retrieval. *Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, May 18-22, 2015, IEEE, Tokyo, Japan, ISBN:978-4-9011-2214-6, pp: 514-517.
- Tursun, O., C. Aker and S. Kalkan, 2017. A large-scale dataset and benchmark for similar trademark retrieval. *Comput. Vision Pattern Recogn.*, 1: 1-33.
- Wei, C.H., Y. Li, W.Y. Chau and C.T. Li, 2009. Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recogn.*, 42: 386-394.
- Yan, Y., J. Ren, Y. Li, J.F.C. Windmill and W. Ijomah *et al.*, 2016. Adaptive fusion of color and spatial features for noise-robust retrieval of colored logo and trademark images. *Multidimension. Syst. Signal Proc.*, 27: 945-968.