EMITTER International Journal of Engineering Technology
Vol. 13, No. 2, December 2025, pp. 173~197
DOI : 10.24003/emitter.v13i2.852

Visual Similarity Detection for Intellectual Property using Deep Transfer Learning

Abeer Al-Nafjan[1], and Mashael Aldayel[2]

[1]Computer Science Department, College of Computer and Information Sciences,
Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
[2]Information Technology Department, College of Computer and Information
Sciences, King Saud University, Riyadh, Saudi Arabia
Corresponding Author: maldayel@ksu.edu.sa

Received October 2, 2025; Revised November 3, 2025; Accepted December 9, 2025

Abstract

Trademarks examination can benefit from deep transfer learning. Utilizing pretrained models to extract image features can significantly improve the trademarks registration process. This approach can facilitate and accelerate image detection. This study aims to enhance the trademark similarity examination process by detecting marks' visual similarities using deep transfer learning. Deep transfer learning has the potential to develop the registration process of trademarks through the implementation of an automated image detection system, which can enhance detection accuracy. To the best of our knowledge, no automated approach has been used locally to determine the similarities between local trademarks. This study proposes an image similarity detection system to make the trademark examination process more efficient and assist examiners in their decision-making. The proposed system was validated using a dataset provided by the Saudi authority for intellectual property (SAIP). To extract the features, we employed a residual network-based convolutional neural network model (ResNet-50). Then principal component analysis (PCA) was used to reduce the number of extracted features. The proposed system reached a mean average precision (MAP) of 0.774, which indicates a promising result in distinguishing the similarity of trademarks. The findings of this research suggested that an image similarity detection system can support decision-making in trademark examination contexts. Trademark examiners, legal professionals, and intellectual property offices can use the results of this research to enhance their evaluation processes and improve the accuracy and efficiency of trademark registration.

Keywords: Deep learning, image processing, trademark, similarity detection.

1. INTRODUCTION

Deep learning is making significant advances in solving problems that have seemed impossible in the past few years, such as facial recognition and self-driving cars [1]. Deep learning models can manipulate various types of data, such as text, images, sounds, and videos, depending on the problem that needs to be solved. The critical feature that distinguishes deep learning from machine learning is feature extraction. In machine learning, features are manually selected and extracted from the input data. In contrast, in deep learning, feature selection and extraction processes are conducted automatically using a neural network with a number of hidden layers [2].

Deep transfer learning can be defined as a regularizer for solving a particular task by passing knowledge from the origin domain to the destination domain using deep learning. The employment of deep learning aimed at simplifying rapid reprocessing of data through the extraction and reengineering of features and employment of transfer learning aimed at improving the knowledge generalization capacity for machine learning [3]. Image recognition has widely used deep transfer learning through the direct use of well-defined pretrained deep learning models such as VGG16, AlexNet, and Inception-v3. Trademarks examination can benefit from deep transfer learning. Using pretrained models to extract image features can improve the Trademarks registration process. This approach can facilitate and accelerate image detection.

The examination and registration process for trademarks poses challenges for examiners. Consider a person who wants to register a new trademark. They should request to register a trademark and fill out the important forms at one of the IP offices. After that, the IP examiners will look at the registrant’s request. To detect any visual similarity, they will upload the trademark image into a dedicated system. The system analyzes each image beforehand to characterize key shape components, categorizing and grouping image regions into families that mirror human image perception to retrieve the closest trademark images that exist in the database based on shape similarity. Thereafter, judgment on whether the mark is similar to an existing one depends on the examiner’s perspective. The examiner will compare the new trademark with a certain number of very similar trademarks recollected from the system.

The motivation of this study is to design an automated system that can assist trademark examiners, especially in managing large databases of marks. In this paper, an image similarity detection system was proposed to aid in the examination process of trademarks. The contributions of this study are as follows: (i) providing a framework that gives examiners a tool to improve the trademark registration process, (ii) building the system based on a deep learning approach with an intuitive interface to help examiners use the system effectively, and (iii) validating the system in collaboration with SAIP using their dataset.

This paper is organized as follows: Section 2 presents the literature review. Section 3 discusses originality of the proposed system. Section 4 describes the system design and methodology. The experiment results and discussion are described in Section 5. Finally, Section 6 concludes with a summary of contributions and future work.

1.1 Background

This section presents a brief explanation of trademarks, deep learning and feature extraction methods. Finally, it describes similarity measurement methods.

1.1.1 Trademarks

A trademark is any recognizable sign that identifies a given enterprise's products and services and distinguishes them from those offered by its competitors [4]. It could be a visible product such as a word, letters, drawings, or symbols, or it could be a non-visible trait such as a scent, color, sound, etc. [5].

In general, trademarks serve two paramount roles for any given product; they identify the product's origin and distinguish it from the others in the same market. Even for small businesses, the importance of logos and trademarks in establishing a brand identity is undeniable. Therefore, most business owners are concerned with creating their figurative trademark and logo that reflects what they provide [6].

A figurative trademark also aims to familiarize the intended consumer with the related product or service. Many intellectual property (IP) offices around the world deal with the registration process. Moreover, a recent rise in trademark applications makes it the most widely used IP right among various sectors. A legally registered trademark awards its owners the exclusive right to sell, license, and use it to develop related businesses [5]. Protecting one's IP involves the ownership of both tangible and intangible assets [4]. Indeed, several reasons lead applicants to register trademarks. First, it is the only way to protect their products and services and prevent imitation by others [6]. Second, it supports the marketing of the company's products. Trademark equity also increases consumer trust and loyalty. Most consumers rely on trademark reputation to make purchasing decisions [7] [8].

Figurative trademark registration is intended to protect legitimate brand owners and consumers. For the brand owner, registration helps ensure that the proposed trademark will not be similar to any registered trademarks; it guarantees that the trademark is a right reserved exclusively by them. Furthermore, it is the only way to legally detect trademark infringement. For consumers, trademarks play an essential role in identifying the origins of products or services. Therefore, using similar trademarks, especially well-known and famous trademarks, can mislead consumers and cause them to confuse brands with one another [6].

1.1.2 Deep Learning

Deep learning is a subfield of machine learning inspired by human cognitive behavior that enables computer systems to recognize reasons for events and learn by example to handle complicated tasks efficiently [9]. The potential of deep learning is represented by how it is used to solve a problem.

Without human intervention, deep learning algorithms design computer models composed of multiple layers until the desired result is obtained based on learning from data [10]. The critical feature that distinguishes deep learning from machine learning is feature extraction. In machine learning, features are selected and extracted from the input data manually. In contrast, in deep learning, the feature selection and extraction processes are conducted automatically using a neural network with a number of hidden layers [11].

Deep neural networks are at the heart of deep learning algorithms and are designed based on knowledge of how the human brain works [9]. Deep neural networks consist of multiple layers, each composed of a collection of units called "neurons" to form a neural network. There are two types of layers: visible and hidden. The visible layers include the input and output layers, and the layers in between are considered hidden layers [12].

There are two fundamental neural network architectures: feed-forward networks and recurrent networks. In a feed-forward network, the connection link between neurons flows in one direction; every node receives input from the preceding layer and produces an output representing the next layer's input without an internal loop. In contrast, a recurrent network feeds the output of a particular node back in as an input of the same node, which makes it similar to brain activities [12]. One of the most commonly used feed-forward networks is the convolutional neural network (CNN).

A CNN is a feed-forward neural network used mostly for image classification and object detection [13]. CNNs have achieved excellent results in the field of image processing over the past decade, where they are involved in the most advanced applications, such as pattern recognition [2]. Furthermore, CNNs require much less preprocessing and achieve more accurate results than other classification techniques [2][12].

A CNN has three main layers: convolutional, pooling, and fully connected layers [10][12]. Typically, the architecture of a CNN takes the form of a sequence of stages. The first stage is composed of two layers: convolutional and pooling. The convolutional layer is the core component of a CNN. It is responsible for performing two tasks: feature extraction and nonlinearity operation [12]. The convolution process is conducted by moving a weighted filter across the input data array and producing a feature map. Low-level features are extracted in an initial layer, and the more important features are extracted in subsequent layers.

Therefore, each layer increases the CNN's complexity while identifying more image features until the desired object is detected [10]. The next layer is the pooling layer, also called a subsampling or down-sampling layer. Typically, it is located between two convolutional layers. This reduces the input received from the preceding layer to produce an output. The purpose of this reduction is to reduce overall network complexity. As in the con-volution process, the pooling operation is conducted by swiping the filter across the input. Instead of a weighted filter, one type of pooling is applied at a time, such as max-pooling [10].

The final layer is fully connected and used to conclude the results. A classification score is calculated in this step using a standard classifier to predict the result's class [9] [12]. Figure 1 shows the basic CNN architecture.

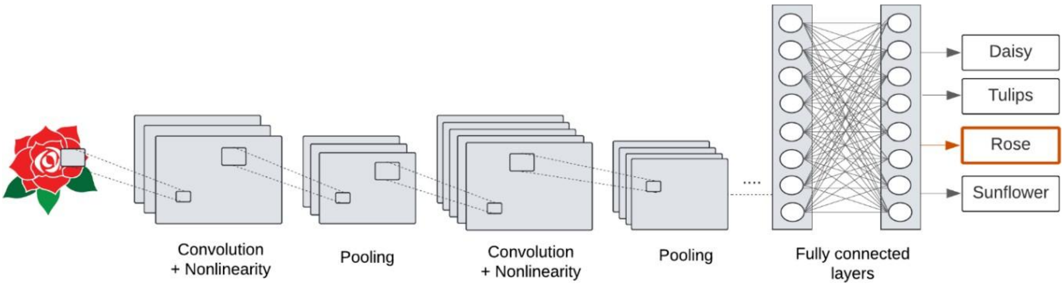


Figure 1. Basic convolutional neural network architecture (CNN)

Different CNN architectures can be used to tackle different problems, but the best one for a given problem can only be found through fine-tuning its parameters using the design of experiments (DOE) method. Indeed, for analogous tasks, numerous CNN architectures that were pre-trained on a large dataset for image classification and achieved excellent accuracy are accessible to reuse.

1.1.3 Deep Transfer learning

Deep transfer learning can be defined as a regularizer for solving a particular task by passing knowledge from the origin domain to the destination domain using deep learning [3]. Image recognition has widely used deep transfer learning through the direct use of well-defined pretrained deep learning models such as VGG16, AlexNet, and Inceptionv3. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [14] was one of the most popular and valuable competitions that encouraged academics and developers to produce such an excellent solution. From 2010 through 2017, an annual competition for object recognition and image classification was held. The ILSVRC encouraged researchers to build robust image processing algorithms by using imageNet [14], a huge, hand-annotated image dataset. The ImageNet dataset contains over 14 million images divided into over 21,000 classes. Many of today's ideas are the result of an ILSVRC competition. Some of the most prominent CNN designs produced during an ILSVRC are AlexNet, VGGNet, and ResNet.

AlexNet: It was introduced in 2012 [15]. AlexNet was trained on over 1.2 million images from the ImageNet dataset to classify them into 1,000 different natural object groups. It has a total of 650,000 neurons and eight layers (5 convolutional layers and 3 fully connected layers). The first, second, and fifth convolutional layers are followed by the max-pooling layer. Furthermore, each neuron’s output is subjected to the rectified linear unit (ReLU). The ReLU function allows AlexNet outperform other activation functions by speeding up the learning process and increasing computational efficiency. The equation of the ReLU function is:

$$f(x) = \max\,(0, x) \tag{1}$$

Data augmentation and dropout approaches were used to reduce overfitting. As a result, AlexNet took first place in the ILSVRC-2012 competition, with top-1 and top-5 test set error rates of 37.5 percent and 17%, respectively.

VGGNet [16]: The architecture of VGGNet was influenced by that of AlexNet. It uses a 3 x 3 kernel-size filter to improve recognition rates over the large-size filter.

VGGNet comes in two different architectures: VGGNet16 and VGGNet19, which have 16 and 19 layers, respectively. Both architectures have three convolutional layers, two max pooling layers, and three fully connected layers. A softmaxis used for classification in the last of the completely connected layers. As a result, VGGNet scored a top-five test set error rate of 6.8%.

ResNet [17]: It was inspired by VGGNet architecture. The main principle behind ResNet is ”Identity shortcut connections,” which mains skip one or more network levels. A 34-layer simple network and a shortcut connection make up the ResNet design. The majority of convolution layers feature three filters and down-sample with a stride of two. The network comes to a close with a global average pooling layer, followed by a fully linked softmax layer. ResNet took first place in the ILSVRC2015 classification competition, with a top-five test set error rate of 3.57 percent.

1.1.4 Image Feature Extraction

Feature extraction is a process that aims to describe the input image information and represent it in a reduced dimensionality containing a set of features. An extracted feature set containing the relevant image information allows for performing the desired task without referring to the full-sized image. The pattern recognition task involves two main steps: feature selection/extraction and classification. The potential accuracy of classification depends on the careful selection of features that help to differentiate each class easily [11].

The extraction method is divided into hand-crafted feature extraction (also called the traditional method) and deep learning feature extraction. The hand-crafted feature extraction method is clarified in the following subsection.

Hand-crafted feature extraction: The most widely used hand-crafted feature extraction methods are color, texture, and shape. Color is one of the leading image features that can be measured and used to differentiate images easily. Moreover, it is not affected by any processing, such as reflection or rotation. A color histogram is an example of the most common method used to describe the color feature for a given image [18].

Texture feature extraction methods describe the intensities or spatial arrangement of the colors in a given image. Lastly, the shape feature is used to describe the image objects using certain spatial characteristics. The most commonly used methods for extracting shape-based features are Hu moments and eccentricity [18].

2. RELATED WORKS

This section presents related works in trademark image retrieval. There are a variety of methods for retrieving similar trademark images such as fuzzy inference system [19], component-based attention [20], vector graphics [21], CNN [22][23], constraint theory [24] and natural language processing (NLP) [7] [25]. Table 1 summarizes the related works in trademark image detection and retrieval systems. The comparison considers the publication date, the method for both feature extraction and classification, dataset, similarity measure and evaluation results.

Authors in [19] proposed trademark image retrieval technique using weighting subtree and the fuzzy inference system. They used tree similarity measurement and the integrate global and local geometric descriptors. The authors evaluated the proposed technique using their own collected 1800 trademark images. The experimental results showed that the proposed technique outperformed other methods, achieving improvements in precision/recall rate of 19.43% and 26.78% for the 416 query images.

Authors in [20] proposed retrieval method of trademark images using component-based attention using middle east technical university (METU) trademark dataset. They extracted features from the components of a trademark image, and then used hard and soft attention mechanisms to weight these features according to their importance. Experimental results showed high mean average precision (MAP) of 25.7 of image retrieval by focusing on the most important components of an image.

Authors in [21] represented trademark images as vector graphics using collected dataset of 1,112 registered trademark images. They extracted the features using angle histogram and relative location of contour segments. They used Euclidean and cosine to measure similarity between retrieved images. Similarity retrieval results showed that their approach outperforms other baseline methods by 13%.

Authors in [22] proposed a method for trademark image retrieval that integrates local binary pattern (LBP) and CNN. They extracted the features using LBP features from 7139 trademark images and METU trademark database, and then used CNN to learn high-level features from these LBP features. They analyzed similarity measures for 7139 trademarks, in another study [24], using constraints theory with pretrained CNN and achieved MAP of 0.68.

Table 1: Summary of trademark image detection and retrieval systems.

Ref	Year	Methods	Dataset	Similarity Measurements	Evaluation Result
[19]	2017	fuzzy inference system and the integration of global and local geometric descriptors.	Collected dataset including MPEG-7	Tree similarity	precision = 95.34
[22]	2017	VGGNet-f and LBP	Collected dataset and METU dataset.	Euclidean distance	Recall =89.63% precision = 45.34%
[23]	2018	ResNet-v2	USPTO dataset	Cosine similarity	MAP = 0.69
[21]	2018	Consider relative location of contour segments	Collected dataset	Euclidean distance and cosine similarity	Similarity retrieval around 100%
[25]	2018	NLP and Histogram algorithm	Not mentioned	Histogram	Not mentioned
[27]	2018	CC, FCC, SIFT, Hu	UK Patent Office	Euclidean distance	Normalized Recall = 0.91
[28]	2018	VGG19	Train: collected trademarks & USPTO. Test: METU	Cosine similarity	Normalized mean rank = 0.046
[24]	2018	CNN along with constraints theory	METU and self-built trademark dataset	Euclidean distance	Recall= 93.74% MAP= 0.68
[26]	2019	Clarifai and Google Vision API services.	Israeli patent office dataset	Clarifai's technology	Recall = 71.3%
[29]	2019	CNN (consists of 13 convolution layers) with U-Net architecture	Training dataset: FlickrsLogos-32 Testing dataset: TopLogos-10	Simple concatenation operation followed by a 1×1 convolution	MAP = 89.2
[30]	2019	CNNs integrated with relevance feedback (CaffeNet)	FlickrLogos-32 and a number of collected trademarks	Euclidean distance	Accuracy=94.14% Precision= 0.973
[20]	2019	Hard and soft attention mechanisms	METU trademark dataset	Euclidean distance	MAP = 25.7
[31]	2019	spatial transformer and recurrent CNN	NPU-TM and METU trademark datasets	Hamming distances	MAP = 0.449
[7]	2020	NLP, Siamese and CNN (VGG-Net16)	Cifar-10 and TopLogo-10 datasets	Euclidean distance	accuracy=97.5%
[32]	2020	VGG16 and image signatures	Collected dataset	Similarity compound formula	MAP= 93.7%
[33]	2021	WTCPN and wavelet transform	Oxford and Holiday dataset	Not mentioned	Accuracy = 91.83%
[34]	2021	instance discrimination and attention mechanism	METU dataset	dot product	NAR= 0.051

Authors in [26] proposed three aspects of trademark similarity: visual, semantic content, and text similarity. They used a combination of Clarifai and Google Vision API services to retrieve similar trademark images using Israeli patent office dataset. Results showed the importance of defining and separating the similarity between trademarks based on these different aspects. They achieved recall success rate of 71.3% using all four similarity aspects.

Authors in [23] used nearest neighbor search and CNN to find the nearest similar image using United States patent and trademark office (USPTO) trademark dataset. Experiment resulted in MAP score of 0.69.

The author in [25] combined visual and conceptual aspects for trademark image retrieval. He extracted text trademark information using NLP and extracted logo features including texture and color using histogram algorithm.

Similarly, Authors in [7] employed NLP along with Siamese and CNN to check the spelling, pronunciation, and image similarity of trademarks using Cifar-10 and TopLogo-10 datasets. Experiments results showed increased similarity accuracy at 97.5%.

Authors in [27] combined different algorithms of feature extraction such as: Concavity/Convexity deficiencies (CC), Scale Invariant Feature Transform (SIFT) Freeman Chain (FC), and Hu Invariant Moments (Hu). They used deep learning and SVM to dynamically detect best results of the four classes that represent the feature extraction algorithms. They used UK Patent Office database of 10,151 images and achieved normalized recall of 0.91 which proved that dynamic selection of extractors can enhance the process of trademarks retrieval.

Authors in [28] proposed trademarks retrieval using combination of CNN and supervised training. They adapted pretrained CNN using two different databases (METU and USPTO) to distinguish visual and conceptual similarities between trademarks. They obtained normalized mean rank of 0.046.

Authors in [29] proposed logo retrieval approach using one-shot learning CNN Experimental evaluations were conducted on benchmark logo datasets: FlickrsLogos32 and TopLogos-10. And achieved MAP of 0.89

Authors in [30] improved trademarks representations of the images using relevance feedback and particle swarm optimization. They used FlickrLogos-32 PLUS database and achieved accuracy of 94.14 % and precision of 0.973. Authors in [31] introduced deep hashing in trademark image retrieval to learn image binary codes using transformation-invariant and recurrent convolutional network. They used two datasets (NPU-TM and METU) and obtained MAP of 0.501.

Authors in [32] proposed an image retrieval system for industrial property area based on deep learning and image processing techniques including image signatures and CNN. The system also used a parallel processing block for dealing with multi-image search scenarios. They used their collected dataset and achieved MAP of 93.7%. Experimental results showed high accuracy and robustness, with linear complexity and processing times compatible with real-time applications.

Authors in [33] proposed retrieval method of artwork image based on wavelet transform and dual propagation neural network (WTCPN). They used two datasets (Oxford and Holiday dataset) and obtained 91.83% accuracy of image retrieval for artwork images.

Authors in [34] introduced unsupervised learning using instance discrimination and lightweight attention network. The experimental results of normalized average rank (NAR) at score of 0.051 on METU dataset showed enhancements compared to most existing supervised learning methods.

As shown in Table 1, there is an efficient and promising potential for the CNN and deep learning techniques to improve image similarity-based retrieval systems. Moreover, the most common metric used to compute the distance between a given query image and the database's images is the Euclidian distance. It is compatible with the nature of trademark images.

3. ORIGINALITY

The first stage of the system analysis is requirements elicitation, concentrating on obtaining the best understanding of users' needs and expectations of such a system with the aim of communicating these needs to system developers. In this study, the requirements were elicited by exploring the current related systems and listing some of the essential requirements based on the problem domain.

An interview was conducted with SAIP examiners to meet their needs. The interview focused on the current system applied by the SAIP examiners to detect trademark similarity, how the registration process is performed, the challenges they face with each query, and how they prevent infringement cases. Moreover, the interview focused on their examination process, the dataset design, and asking for permission to access the data for system development purposes.

In regard to the examination process, consider a person who wants to register a new trademark. They should submit a request to register a trademark and fill out the required forms at one of the IP offices. After that, the IP examiners will look at the registrant's request. To detect any visual similarity, they will upload the trademark image into a dedicated system. The system analyzes each image beforehand to characterize key shape components, categorizing and grouping image regions into families that mirror human image perception to retrieve the closest trademark images that exist in the database based on shape similarity. Thereafter, the judgment of whether the mark is similar to an existing one depends on the examiner's perspective. The examiner will compare the new trademark against a certain number of very similar trademarks recollected from the system.

Regarding the dataset, Nice classification (NCL) is an international classification for goods and services, which is applied for the purposes of registering trademarks [35] [36]. It was established in 1957 by the Nice Agreement. NCL consists of 45 classes, the first 34 classes were for goods and the remaining were for services. Hence, the IP offices of the contracting states must follow this classification and indicate, in the official registration documents, the class in which the good or service belongs [35]. Therefore, each registered trademark on the SAIP belongs to one of the 45 classes.

infringement cases include existing visual or semantic similarities. One of the study's objectives was to help IP examiners detect visual trademark similarities to prevent infringement cases automatically. The purpose of the study was to develop a system that can discover and recall trademarks that are registered and similar to an input query image automatically and accurately. System users include individual users and SAIP examiners in Saudi Arabia. The system aims to provide assessments to IP examiners as they conduct examinations by retrieving the most similar trademark images to the input query. In addition, it allows individual users to ensure their trademark is exclusive to them and is not similar to another registered trademark.

4. SYSTEM DESIGN

Image retrieval is the process of searching for and identifying images stored in the database that contain similar visual content to a given query. With the increasing number of images worldwide, image retrieval operations play an essential role in many fields, such as medical imaging, e-commerce applications, and autonomous driving tasks [37]. Figure 2 shows a block diagram of trademark similarity detection system.

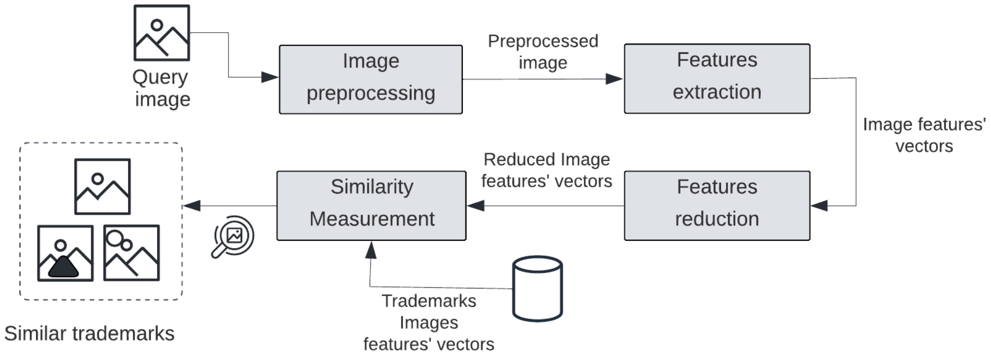


Figure 2. Block diagram of trademark similarity detection system

To summarize, any trademark similarity detection system consists of the following processing stages: image processing, features extraction, features reduction, and similarity measurement.

The proposed system operation (as shown in Figure 3) involves image feature extraction using ResNet-50, feature reduction using PCA, Similarity measurement using Euclidian distance. This study used two datasets: Middle East Technical University (METU) and Saudi authority for intellectual property (SAIP) datasets. In this recent study [36], deep learning techniques are used to automatically extract image features to retrieve trademarks based on similarity in shape using METU dataset. Their findings indicated that the ResNet50 architecture achieved superior results compared to the VGG-16 model. Building upon this research, the current work focuses on detecting visual similarity and extracting features within the SAIP dataset using the ResNet-50 architecture.

The following subsections provide an illustration of the methodology employed by the proposed system, along with details regarding its implementation. The benchmark dataset and the pre-processing steps undertaken are first described. A description of the extracted features and the employed similarity measurement technique is then presented. Finally, a discussion on the evaluation method is included.

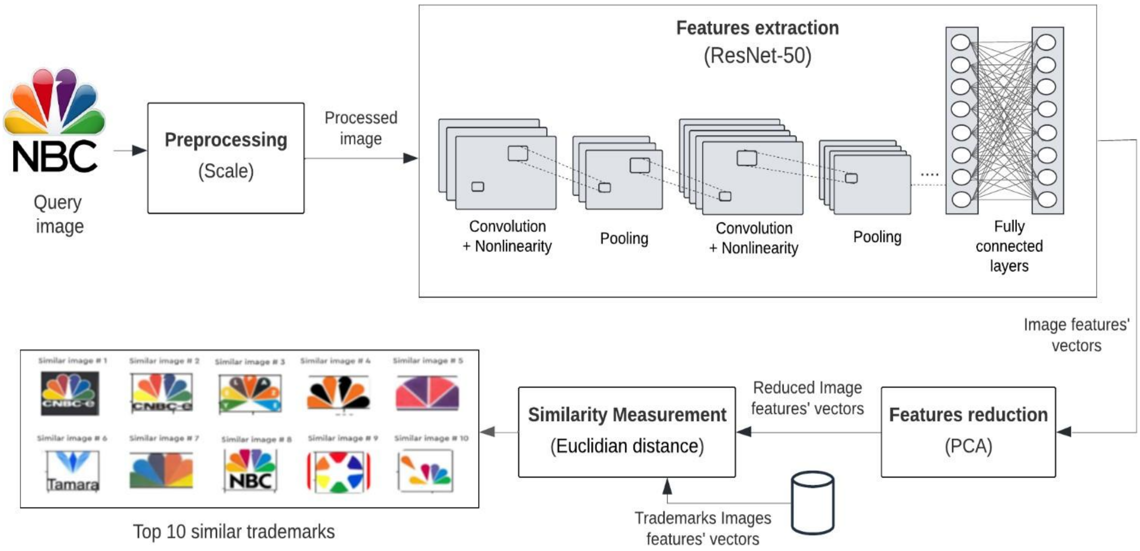


Figure 3. The proposed System architecture

4.1 Dataset Description

This study used two datasets: METU and SAIP datasets which are clarified in the following subsections.

METU dataset: In 2014, Tursun and Kalkan [38] introduced the large trademark dataset called METU, which contains 923,343 images of trademark that belong to about 410,000 businesses. METU images can be categorized into shape, text, or a mixture of them based on their contents. METU is a valuable dataset in training models of trademark detection and retrieval systems.

METU has been used for conducting a number of studies, and it has been proved that it is well suited for testing new algorithms [24] [20] [31]. METU is available upon request for researchers in related fields. An example of the METU images that belong to the same class is shown in Figure 4.

SAIP dataset: SAIP is the national authority that aims to regulate, support, protect, and promote the fields of IP in Saudi Arabia according to international best practices [39].

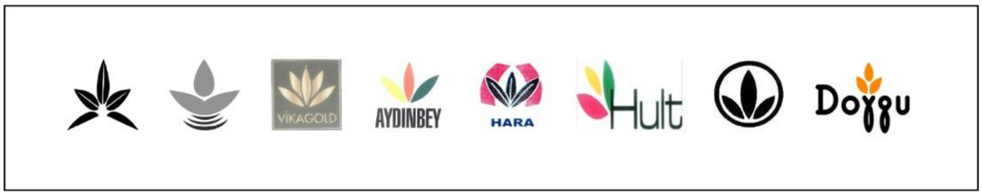


Figure 4. Example of the METU images that belong to the same class.

To achieve this goal, the organization undertakes many tasks, such as proposing and developing related IP laws and regulations, registering and protecting IP rights, and sharing IP rights information with the public. Moreover, it is considered an official body representing the Kingdom of Saudi Arabia (KSA) in international and regional IP communities. Any new trademark should be examined to ensure that it is not similar to registered marks in the same sector. There are two main types of marks: textual and figurative. SAIP examiners review a textual trademark only to ensure no semantic similarities among those registered, without considering other parameters such as font type or style. In contrast, the figurative trademark should not be visually similar to any other previously registered mark. Note that the similarity judgment between trademarks depends on the examiner's subject knowledge and experience, which can vary from one examiner to another. Once the trademark is registered, it will be protected for 10 years with the possibility of extension.

In collaborating with the SAIP, one thousand trademarks were used to verify the retrieval performance, involving locally registered and well-known international marks. The trademarks belong to 10 different classes. Note that most of the dataset trademarks are text only. Figure 5 shows the SAIP samples.

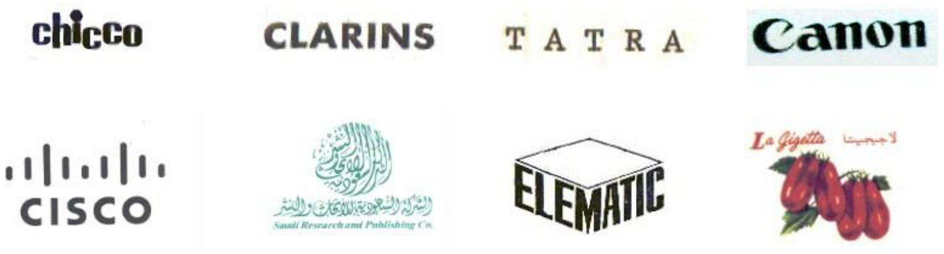


Figure 5. The SAIP dataset samples.

4.2 Data Preprocessing

Each deep neural network requires special input characteristics to operate effectively, such as standardized image size. Moreover, the performance of a deep neural network depends mainly on its capability and the abundance of training data. Therefore, some processing should be conducted on the individual image or set of images before moving forward into the network.

Resizing: The first thing that should be processed is the size of the images. Each CNN, especially if pre-trained, requires a specific size for all input data. The resizing process can be conducted using external tools or by applying a piece of code to the individual images or the entire dataset. Each image was resized and padded into a 224 × 224 image, which is the size required by ResNet-50.

Data Augmentation: A process that aims to artificially increase the number of dataset samples is called data augmentation [40]. It is conducted by making a series of random changes to the original training set to generate similar but different training samples. Data augmentation includes several techniques, such as cropping, flipping, rotating, zooming in or out, and changing image colors. The main aim of this type of augmentation is to improve the dataset for the training process.

The character of trademark images is artificial that requires little or no processing. This study mainly focused on visual similarity among trademarks. Because the images in the dataset are a variety of sizes, all images were re-scaled to the pre-defined element by a given CNN. Furthermore, since the dataset is large, it did not require any augmentation.

4.3 Feature Extraction

Feature extraction is a process that aims to describe the input image information and represent it in a reduced dimensionality containing a set of features. An extracted feature set containing the relevant image information allows for performing the desired task without referring to the full-sized image. The extraction method is divided into hand-crafted feature extraction (also called the traditional method) and deep learning feature extraction. Nowadays, hand-crafted feature extraction is considered time-consuming compared to more recent techniques. Therefore, deep learning feature extraction was used in this study.

Deep learning feature extraction: A wide range of image-related intelligent tasks have been achieved by extracting image features efficiently and automatically using deep learning. In recent years, most image processing systems have relied on CNNs to extract image features within a certain timeframe. Low-level features such as edges, corners, and texture are extracted in the first layers of the CNN, and then these features are combined toward learning high-level features, such as objects [12]. The depth of the CNN depends on the problem being solved. The output of the CNN will be feature vectors usable for the classification step or any further process.

Automating the image-similarity-detection operation involves extracting image features without human intervention. Therefore, a deep learning application could improve related system results due to its ability to extract deep features, such as detecting an intended object in a given image efficiently within a short time frame.

Thus, ResNet-50, a pre-trained CNN, was utilized to extract the dataset image features. It comprised of 16 residual blocks, each containing multiple convolutional layers stacked with bypass connections known as residual connections. This enables ResNet-50 to effectively learn complex feature representations even with its significant depth (50 layers) [17].

Features are extracted from the output of each individual convolutional layer within the network, resulting in a with the final layer potentially reaching 2048 dimensions. Furthermore, image features were obtained by the last pooling layer before the fully connected layer, which was not included, with an output dimension of 2048. CNN processed both query and dataset images, in which the dataset images were stored in the database as vectors.

4.4 Feature Reduction

The techniques utilized to reduce the size of the vectors are referred to as feature reduction. ResNet-50's features are excessively large in size; consequently, reducing the size of the features is necessary to avoid future problems.

Principal component analysis (PCA) is an unsupervised technique for reducing dimensionality in machine learning [41]. In order to increase detection performance, PCA was applied to reduce the extracted features. Features scaling was unnecessary because the features had already been normalized. Furthermore, the features extracted by the ResNet-50 represented 1,450 components. These components explained more than 95% of the variance.

4.5 Similarity measurement

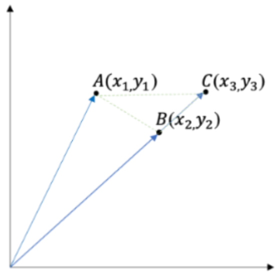


Figure 6. Representation of vector/feature space.

Similarity measurement defines how much two objects are alike and is measured by calculating the distance between them [42]. Consider three objects, A, B, and C, located in the same vector/feature space as shown in Figure 6. Each object in the form of a vector represents an image. If the distance between two vectors is too small, these images are similar to each other.

Similarity measurement is considered a primary component of any similarity-based retrieval system. There are many ways to compute the distance between two vectors. The most popular similarity measurement methods are described below.

Euclidean distance is the most common distance measure. It computes the length of the path or segment between two points, such as A and B. Practically, the Pythagorean theorem can be used to calculate Euclidean distance. The mathematical equation is given below [13]:

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (2)$$

Cosine similarity is a metric used to measure the cosine of the angle between two given vectors in the same vector space [42]. In some cases, cosine similarity is not effective on its own when the two vectors are collinear, such as B and C in figure 6. Below is the cosine similarity equation:

$$\text{Sim(A,B)} = \cos \llbracket(\theta) = (A \,.\, B)/\|A\|\|B\| \rrbracket \qquad (3)$$

Manhattan distance (Mdist) measures the distance between two vectors by calculating the total sum of the difference between the x-coordinates and y-coordinates. The Mdist equation is given below [13]:

$$\text{Mdist} = \sum_{i=1}^{k} |x_i - y_i| \qquad (4)$$

Hamming distance is the metric used to compute the number of attributes on which the two vectors differ. Thus, the Hamming distance between two points u and v is the number of places in which uandv differ[13]. In other words, the Hamming distance equation is given below [43]:

$$\text{Hamming Distance} = \sum_{\mathrm{i}=1}^{\mathrm{n}} |\mathrm{u_i} - \mathrm{v_i}| \qquad (5)$$

In this study, the output of the feature reduction module was used as the feature vector. Thus, to compute the similarity between the query and the stored dataset images, some computation must be conducted to calculate the distance between their feature vectors.

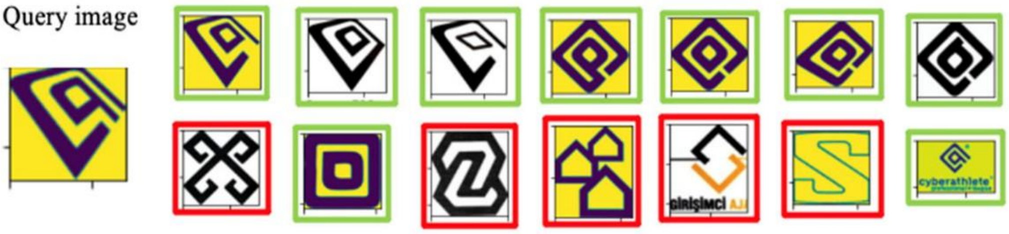


Figure 7. Ranked images retrieved for a given query

In this study, the Euclidian distance method was used. The distance was computed from the query image to each image in the database. Next, the images were ranked in increasing order based on the distance. Then, the system retrieved n ranked images corresponding to the user input (as shown in Figure 7).

Consider the feature vector q for a given query and set of database features vectors a_1, a_2,..., a_n. The system measures the similarity between q and every feature vector from a_1 to an. The Euclidian distance is then calculated using the following equation [13]:

$$d(a_i, q) = (\sum_{i=1}^{n} |a_i - q|^2)^{\frac{1}{2}} \qquad (6)$$

4.6 Graphical User Interface

Figure 8 illustrates the proposed system's main workflow. Practically, when new trademark application submitted. The trademark examiner will upload the image using the proposed system. The system will compare the image to the registered trademarks images and display similar trademarks, ranked by similarity.

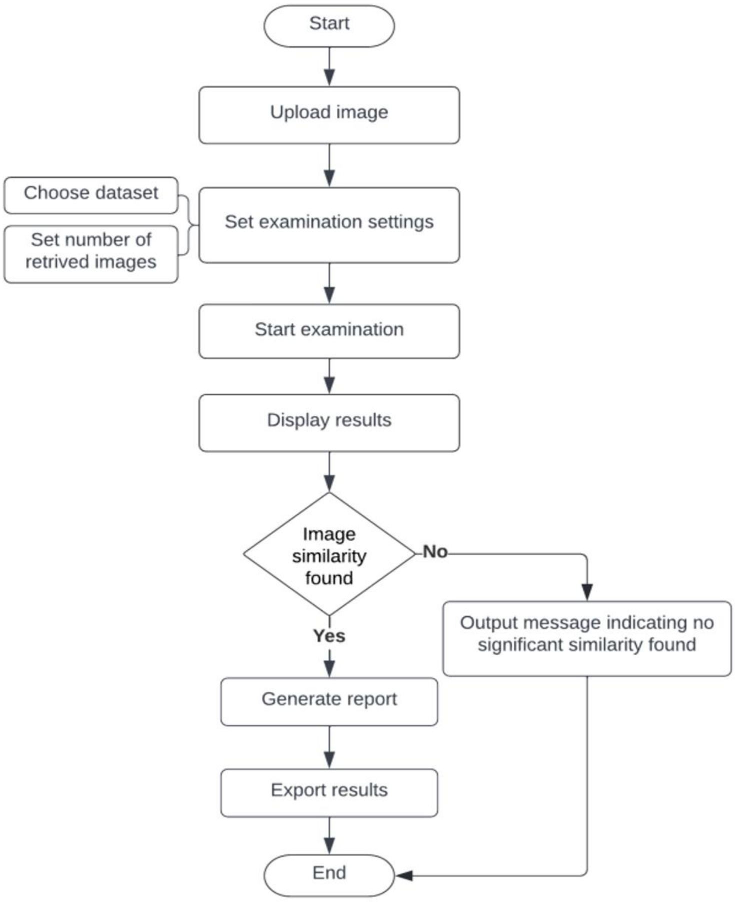
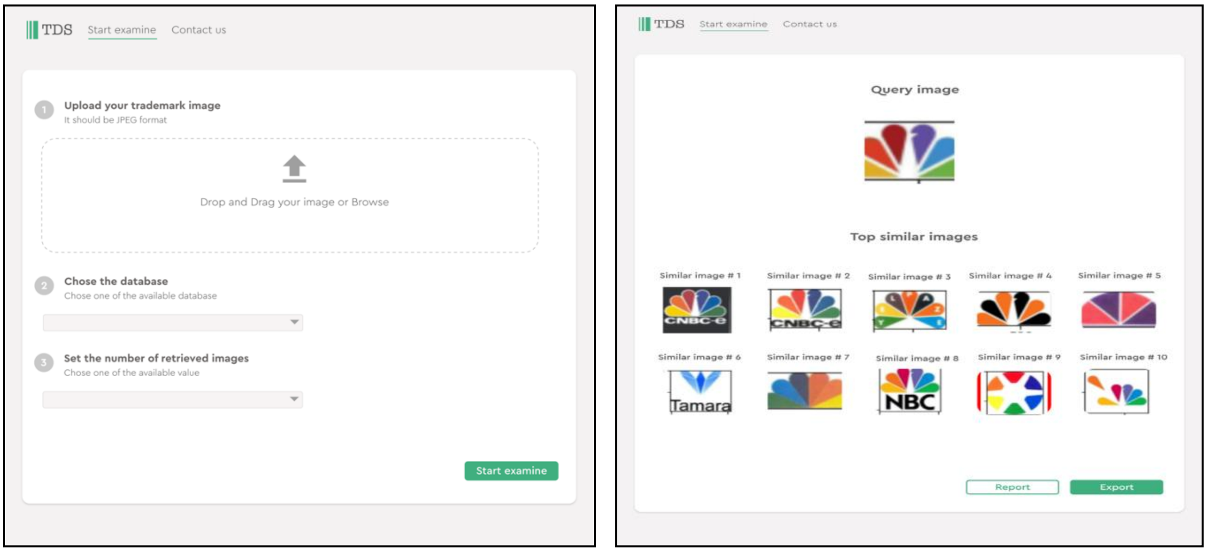


Figure 8. Flow chart of the proposed system.

This system was supported by an intuitive graphical user interface (GUI) to facilitate interaction with the system. The prototype of the GUI was designed with the help of Adobe XD. Figure 9 presents the primary pages of the system GUI.

5. EXPERIMENT AND ANALYSIS

For classification and clarification, trademarks should be designed to be clearly distinguishable from each other. Therefore, the given trademarks were considered marks of the same class. The model was trained on METU dataset which contains 923,343 images. Then validated on SAIP dataset with one-thousand of the trademarks, involving locally registered and well-known international marks.



1 Upload your trademark image
It should be JPEG format

Drop and Drag your image or Browse

2 Chose the database
Chose one of the available database

3 Set the number of retrieved images
Chose one of the available value

Query image

Top similar images

(a) (b)

Figure 9. GUI of the system.

We have evaluated the system performance in study [36] using METU dataset and achieved a MAP of 0.774, indicating good retrieval accuracy for trademark images, including both locally registered and well-known international marks. MAP was calculated as the average precision across all trademark classes in the dataset.

To validate the system performance on the SAIP dataset, this section explains how we used query set of four images of trademark. The query set contained four images, one of which was a trademark rejected by the SAIP examiners because it was similar to a registered mark. In addition, three other trademarks were created to simulate infringement cases. Therefore, each query had at least one of the relevant marks that should be detected as similar. Figure 10 shows the original and imitated/rejected trademarks.

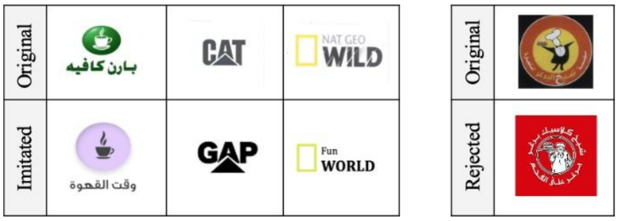


Figure 10. Original and imitated/rejected trademarks for SAIP case study.

As a result, the relevant trademark of each query appeared in the top 10 marks retrieved by the system. That means the proposed approach gives a promising result in helping the IP examiners retrieve the most similar trademark to a given query. It should be noted, however, that the similarity judgment between given trademarks primarily depends on the examiner's experience.

The most shared dataset images contained only text. The system could not detect semantic similarities accurately where the word is considered as a shape. Therefore, fonts affect the detection operation. Figure 11 presents the results of the top retrieved images (until the correct similar image) regarding each query.

Query (a) was rejected by the SAIP examiners due mainly to a semantic similarity with a registered mark. At the same time, it was visually similar, too. It had a circular shape; therefore, most of the retrieved images were circles surrounded by frames or texts.

Query (b) could be considered a shape with three objects separated by spaces. Hence, the system retrieved the correct similar mark beside the marks with the same arrangement or shape. Query (c) contained a clear shape; therefore, the system was able to retrieve the most similar images correctly. The last query (d) contained text, which misled the system. Therefore, the detection performance was affected by each letter.

The suggested model was compared to the previous studies [38] [44] [34] that used same METU dataset (Table 2).

Table 2. Comparison of evaluation results with previous research

Ref	Methods	NAR Result
[38]	Hard and soft attention mechanisms + VGG16	0.086
[34]	instance discrimination and attention mechanism + ResNet50	0.051
[44]	Canny algorithm to extract edges, then split into components, then aggregate with Fisher Vector	0.083
Proposed model	feature extraction using ResNet-50	0.073

Authors in [38] [20] used VGG16 based CNN and achieved a NAR of 0.086, authors in [44] achieved a NAR of 0.083, and authors in [34] used ResNet50 model and achieved a NAR of 0.110. The proposed model got the best results with an average rank of 67,067.788 , NAR of 0.073, and MAP of 0.774. Furthermore, 85 percent of the images related to a specific query were among the first 20 retrieved images.

6. CONCLUSION

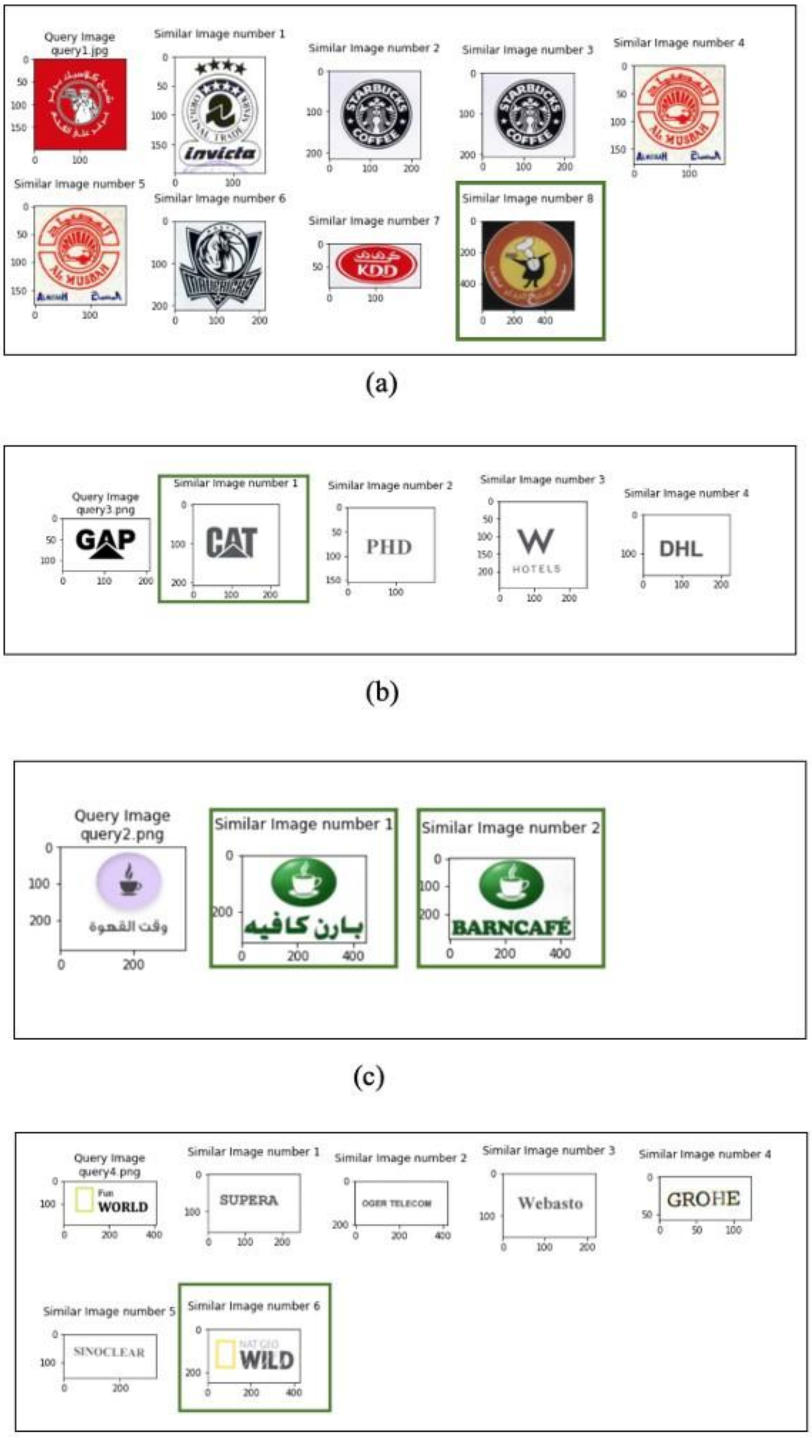
The contribution of this work is to provide a framework and comprehensive background for trademark similarity experiments as an examination tool in the trademark registration process. The following are the main contributions: (i) proposing a framework for a trademark similarity detection system; this aimed to design and develop a system that reads an image, extracts its features, and finds the similarity between a given image and each of those registered and stored based on their vector distance. The system architecture consists of four main modules: preprocessing, feature extraction, feature reduction, and similarity measurement.

(ii) building the system based on deep learning; the system was built based on pre-trained CNNs to extract image features using ResNet-50 as a feature extractor. (iii) validating the proposed system with the collaboration of the SAIP to test the effectiveness of the proposed system. One thousand trademarks were used to verify the retrieval performance, involving both locally registered and well-known international marks.

The system was validated using the SAIP dataset. As a result, the system was able to retrieve similar images about a query that had a clear shape. In contrast, the image containing text could not be detected accurately, and the detection operation was affected by the shape of each letter and font type.

The majority of SAIP dataset pictures were text-only. The proposed system was unable to detect semantic similarities accurately, thereby the word was considered as a shape. Thus, in future, the system can be improved by consider semantic similarities among text-only trademarks. Moreover, to improve its performance, the ResNet-50-based CNN model will be fine-tuned with a larger annotated trademarks dataset.

A collaboration between the researcher and domain experts is necessary to design and build a beneficial trademark similarity detection system. AI researchers, developers, and trademark experts, domain experts, particularly, classification experts that are manually working trademark image classification and trademark examiners who are manually searching similar images. More research contributions are needed in artificial intelligence algorithms and computational methods to design and build more complex architectures and to explore the performance of deep learning models with larger image sizes and higher resolutions which will minimize data loss.



(a)

(b)

(c)

(d)

Figure 11. The retrieved images for each query.

Acknowledgments

Our thanks go out to the staff of SAIP for their help in offering us the dataset and resources to evaluate the system.

REFERENCES

[1] S. Russell and P. Norvig, **Artificial Intelligence A Modern Approach**, *Pearson (Hoboken)*, Ed. 4, 2020.

[2] S. Albawi, T. A. Mohammed, and S. Al-Zawi, **Understanding of a convolutional neural network**, *Proceedings of the 2017 International Conference on Engineering and Technology (ICET)*, Antalya, pp. 1-6, 2017.

[3] X. Gu *et al.*, **EEG based Brain-Computer Interfaces (BCIs): A Survey of Recent Studies on Signal Sensing Technologies and Computational Intelligence Approaches and their Applications, Signal Processing and Machine Learning for Brain-Machine Interfaces**, pp. 1-22, 2020.

[4] F. Mohd Anuar, R. Setchi, and Y. K. Lai, **Trademark image retrieval using an integrated shape descriptor**, *Expert Systems with Applications*, Vol. 40, No. 1, pp. 105-121, 2013.

[5] C. V. Trappey, A. J. C. Trappey, and B. H. Liu, **Identify trademark legal case precedents - Using machine learning to enable semantic analysis of judgments**, *World Patent Information*, Vol. 62, pp. 101980, 2020.

[6] S. Y. Arafat, M. Saleem, and S. A. Hussain, **Comparative analysis of invariant schemes for logo classification**, *Proceedings of the 2009 International Conference on Emerging Technologies,* Islamabad, pp. 256-261, 2009.

[7] C. V. Trappey, A. J. C. Trappey, and S. C. C. Lin, **Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies**, *Advanced Engineering Informatics*, Vol. 45, pp. 101120, 2020.

[8] W. Macías and J. Cerviño, **Trademark dilution and its practical effect on purchase decision**, *Spanish Journal of Marketing - ESIC*, Vol. 21, No. 1, pp. 1-13, 2017.

[9] Y. LeCun, Y. Bengio, and G. Hinton, **Deep learning**, *Nature*, Vol. 521, No. 7553, pp. 436-444, 2015.

[10] C. C. Aggarwal, **Neural Networks and Deep Learning**, *Springer International Publishing (Cham)*, Ed. 1, 2023.

[11] G. Kumar and P. K. Bhatia, **A Detailed Review of Feature Extraction in Image Processing Systems**, *Proceedings of the 2014 Fourth International Conference on Advanced Computing Communication Technologies*, Rohtak, pp. 5-12, 2014.

[12] N. Singh and H. Sabrol, **Convolutional Neural Networks: An Extensive arena of Deep Learning. A Comprehensive Study**, *Archives of Computational Methods in Engineering*, Vol. 28, No. 7, pp. 4755-4780, 2021.

[13] S. Russell and P. Norvig, **Artificial Intelligence A Modern Approach**, *Pearson Education, Inc. (Hoboken)*, Ed. 4, 2010.

[14] O. Russakovsky *et al.*, **ImageNet Large Scale Visual Recognition Challenge**, *International Journal of Computer Vision (IJCV)*, Vol. 115, No. 3, pp. 211-252, 2015.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, **ImageNet Classification with Deep Convolutional Neural Networks**, *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1097-1105, 2012.

[16] K. Simonyan and A. Zisserman, **Very deep convolutional networks for large-scale image recognition**, *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, pp. 1-14, 2014.

[17] K. He, X. Zhang, R. S. Ren, and J. Sun, **Deep Residual Learning for Image Recognition**, *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, pp. 770-778, 2016.

[18] A. J. C. Trappey, C. V. Trappey, and S. C. C. Lin, **Detecting Trademark Image Infringement Using Convolutional Neural Networks**, *Proceedings of the 26th ISTE International Conference on Transdisciplinary Engineering*, Tokyo, pp. 477-486, 2019.

[19] C. S. Chen and C. M. Weng, **An Efficient Retrieval Technique for Trademarks Based on the Fuzzy Inference System**, *Applied Sciences*, Vol. 7, No. 8, pp. 849, 2017.

[20] O. Tursun *et al.*, **Component-Based Attention for Large-Scale Trademark Retrieval**, *IEEE Transactions on Information Forensics and Security*, Vol. 17, pp. 2350-2363, 2022.

[21] K. Abe, H. Morita, and T. Hayashi, **Similarity Retrieval of Trademark Images by Vector Graphics Based on Shape Characteristics of Components**, *Proceedings of the 2018 10th International Conference on Computer and Automation Engineering*, New York, pp. 826, 2018.

[22] T. Lan, X. Feng, Z. Xia, S. Pan, and J. Peng, **Similar Trademark Image Retrieval Integrating LBP and Convolutional Neural Network**, *Proceedings of the 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC),* Shenzhen, pp. 231-242, 2017.

[23] G. Showkatramani, S. Nareddi, C. Doninger, G. Gabel, and A. Krishna, **Trademark image similarity search,** *Communications in Computer and Information Science*, Vol. 850, pp. 199-205, 2018.

[24] T. Lan, X. Feng, L. Li, and Z. Xia, **Similar Trademark Image Retrieval Based on Convolutional Neural Network and Constraint Theory**, *Proceedings of the 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, Xi'an, pp. 1-6, 2018.

[25] S. Khamkar, **User Retrieval of Trademarks System Using Conceptual Similarity Approach**, *HELIX*, Vol. 8, No. 5, pp. 3754-3758, 2018.

[26] I. Mosseri, M. Rusanovsky, and G. Oren, **TradeMarker - Artificial Intelligence Based Trademarks Similarity Search Engine**, *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, pp. 97-105, 2019.

[27] S. B. K. Aires, C. O. A. Freitas, and M. L. Sguario, **Dynamic Selection Feature Extractor for Trademark Retrieval**, *Computational Science and Its Applications – ICCSA 2018*, pp. 219-231, 2018.

[28] C. A. Perez *et al.*, **Trademark Image Retrieval Using a Combination of Deep Convolutional Neural Networks**, *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, pp. 1-7, 2018.

[29] A. K. Bhunia, A. K. Bhunia, S. Ghose, A. Das, P. P. Roy, and U. Pal, **A deep one-shot network for query-based logo retrieval**, *Pattern Recognition*, Vol. 96, pp. 106965, 2019.

[30] H. M. Y, S. A, and Jaber, **Trademark image recognition utilizing deep convolutional neural network**, *Annals of Electrical and Electronic Engineering*, Vol. 2, No. 3, pp. 13-20, 2019.

[31] Z. Xia, J. Lin, and X. Feng, **Trademark image retrieval via transformation-invariant deep hashing**, *Journal of Visual Communication and Image Representation*, Vol. 59, pp. 108-116, 2019.

[32] S. Jardim, J. António, C. Mora, and A. Almeida, **A Novel Trademark Image Retrieval System Based on Multi-Feature Extraction and Deep Networks**, *Journal of Imaging*, Vol. 8, No. 9, pp. 238, 2022.

[33] J. Wan and Y. Xiaobo, **Intelligent Retrieval Method of Approximate Painting in Digital Art Field**, *Scientific Programming*, Vol. 2021, pp. 1-8, 2021.

[34] J. Cao, Y. Huang, Q. Dai, and W. K. Ling, **Unsupervised Trademark Retrieval Method Based on Attention Mechanism**, *Sensors*, Vol. 21, No. 5, pp. 1894, 2021.

[35] World Intellectual Property Organization, **Nice Agreement Concerning the International Classification of Goods and Services for the Purposes of the Registration of Marks**, *WIPO (Geneva)*, 2009.

[36] H. Alshowaish, Y. Al-Ohali, and A. Al-Nafjan, **Trademark Image Similarity Detection Using Convolutional Neural Network**, *Applied Sciences*, Vol. 12, No. 3, 2022.

[37] S. Gkelios, A. Sophokleous, S. Plakias, Y. Boutalis, and S. A. Chatzichristofis, **Deep convolutional features for image retrieval**, *Expert Syst. Appl.*, Vol. 177, pp. 114940, 2021.

[38] O. Tursun and S. Kalkan, **METU dataset: A big dataset for benchmarking trademark retrieval**, *Proceedings of the 2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, Tokyo, pp. 514-517, 2015.

[39] Saudi Authority for Intellectual Property (SAIP), **About the Saudi Authority for Intellectual Property**, 2022.

[40] J. M. Czum, **Dive Into Deep Learning**, *Journal of the American College of Radiology*, Vol. 17, No. 5, pp. 637-638, 2020.

[41] A. Tharwat, **Principal component analysis - a tutorial**, *International Journal of Applied Pattern Recognition*, Vol. 3, No. 3, pp. 197, 2016.

[42] L. Deng, **Deep Learning: Methods and Applications**, *Foundations and Trends in Signal Processing*, Vol. 7, No. 3-4, pp. 197-387, 2014.

[43] J. K. Min, R. T. Ng, and K. Shim, **Efficient Aggregation Processing in the Presence of Duplicately Detected Objects in WSNs**, *Journal of Sensors*, Vol. 2019, pp. 1-15, 2019.

[44] Y. Feng, C. Shi, C. Qi, J. Xu, B. Xiao, and C. Wang, **Aggregation of reversal invariant features from edge images for large-scale trademark retrieval**, *Proceedings of the 2018 4th International Conference on Control, Automation and Robotics (ICCAR),* Auckland, pp. 384-388, 2018.