Trademark Image Similarity Search

Girish Showkatramani, Sashi Nareddi, Chris Doninger, Greg Gabel, and Arthi Krishna(✉)

United States Patent and Trademark Office, Alexandria, VA, USA
{Girish.Showkatramani,Sashi.Nareddi,Chris.Doninger, Greg.Gabel,Arthi.Krishna}@uspto.gov

Abstract. A trademark may be a word, phrase, symbol, sound, color, scent or design, or a combination of these, that identifies and distinguishes the products or services of a particular source from those of others. One of the crucial steps both prior to filing of the trademark applications as well as during the review of these applications is conducting a thorough trademark search to determine whether the proposed mark is likely to cause confusion with prior registered trademarks and pending trademark applications. Currently, the trademark applicants or their representatives and examining attorneys manually search the United States Patent and Trademark Office (USPTO) database that contains all of the active and inactive trademark registrations and applications. This search process relies on words and Trademark Design codes (which are hand annotated labels of design features) to search for images, thereby limiting the overall search process to primarily text-based search. For marks having image characteristics, users visually look at the image and other design characteristics and compare it with existing registered or pending trademarks to determine its uniqueness. Overall, the process of exhaustively looking at all the images that are categorized using a specific design code, while comprehensive, may take a substantial amount of time.

Recently, Convolutional Networks (CNNs) have revolutionized the field of computer vision and demonstrated excellent performance in image classification and feature extraction. In this study, we utilize CNN to address the problem of searching trademarks similar to a chosen mark based on the image characteristics. A corpus of trademark images are pre-processed and then passed through a trained neural network to extract the image features. We then use these features to perform image search using the approximate nearest neighbor (ANN) variant of the nearest neighbor search (NNS) algorithm as depicted in Fig. 2. NNS is a form of proximity search that aims to find closest (or most similar) data points/items from a collection of data points/items.

This system thereby seeks to provide an efficient image-based search alternate to the current keyword and category of design code combination of searching.

Keywords: Approximate nearest neighbors · CBIR · Deep learning LIRE · Solr · Trademark image search

C. Stephanidis (Ed.): HCII Posters 2018, CCIS 850, pp. 199–205, 2018.
https://doi.org/10.1007/978-3-319-92270-6_27

1 Introduction

A trademark may be a recognizable word, phrase, symbol, sound, color, scent or design, or a combination of these, that identifies products or services of an individual, an organization or a particular source from those of others. The two primary purposes of a trademark are to: (1) protect brand names and logos used on goods and services and give the trademark owner the exclusive right to use the mark; and (2) act as a source indicator for consumers to ensure that the products and services they utilize under particular brands emanate from the sources that they expect [1]. Selecting a mark is the first step in the overall trademark application/registration process. One of the key factors in choosing a mark and filing it for registration is determining whether a "likelihood of confusion" [2] exists with the mark that is being filed or anything that has already been registered or filed. USPTO examines every trademark application for compliance with federal rules and laws and grants registrations when, among a host of other factors, no likelihood of confusion exists. In fact, "likelihood of confusion" between the mark that is being filed and a mark already registered or in a pending application, is the most common reason for refusal of a trademark application. Therefore, before the trademark filing process, each trademark applicant is strongly encouraged, though not required, to conduct a thorough trademark search to determine whether the proposed mark is likely to cause confusion with any existing registered trademarks or pending trademark applications [1].

Currently, the trademark applicants, and/or their attorneys and representatives, manually search the USPTO's database of active and inactive trademark registrations and applications using the Trademark Electronic Search System (TESS) search engine. This search engine provides access to crucial information such as text and images of registered marks, and marks in pending and abandoned applications. During this search phase, trademark applicants, or their attorneys or representatives, visually identify and determine whether there are any same or similar marks for related goods and/or services that have already registered or are pending. Furthermore, a thorough study of each mark is required to determine that the goods and services are not related [1]. In addition, once the trademark application has been filed, it is forwarded to a trademark-examining attorney for legal review. During the review phase, the USPTO examining attorneys also manually search existing USPTO records of registered trademarks and prior pending applications to determine potential likelihood of confusion using the USPTO search system (known as X-search, which utilizes the same database as TESS, though the interface differs). Overall, the process of manually researching and identifying marks with similar text and image characteristics is a complex task that often takes a substantial amount of time.

Recently, Content-Based Image Retrieval (CBIR) systems have led to advancement in image retrieval and recognition methods by finding and retrieving images independent from the metadata. In CBIR, image global and local low-level features are extracted by their visual content such as shape, texture, and color or any other information that can be derived from the image itself. Similarly, Convolutional Networks (CNNs) [3] have achieved great success in the field of computer vision and demonstrated excellent performance in large-scale image classification [4] and object detection [5]. Moreover, in the last few years, CNNs have emerged as a methodology in extracting features [6, 7] such as basic shapes, textures, and colors etc. from the unlabeled data. Most notably, a significant advancement in the deep learning-based methods has been seen after Krizhevsky et al. [8] achieved the first place on the ILSVRC 2012 challenge using a CNN model that achieved top 1 and top 5 error rates of 37.5% and 17.5%. This has been made possible due to the rapid growth in the amount of annotated data [9], powerful graphic processing units (GPUs) [10] and advancements in computing architecture. Additionally, in the last few years, the depth of CNNs has advanced greatly from 8 layers (AlexNet) [8] to 19 layers (VGGNet) [4], 22 layers (GoogleNet) [11], and even 152 layers (ResNet) [12], improving the overall classification accuracy. Furthermore, numerous deep learning libraries and platforms such as TensorFlow [13], Theano [14], Caffe [15], Torch [16], Computational Network Toolkit [17] etc. have been developed and made available in the open source platform, enabling further research in simplifying the complexity of deep neural networks.

In this paper, we address the problem of searching trademarks similar to a chosen mark using a neural network pre-trained on the trademark dataset. TensorFlow-Slim high level neural network API library was firstly used to extract the image features from the pre-trained Inception-ResNet-v2 [18] neural network. The approximate nearest neighbor algorithm was then used to identify the "nearest neighbors", that is, trademarks similar to the input mark.

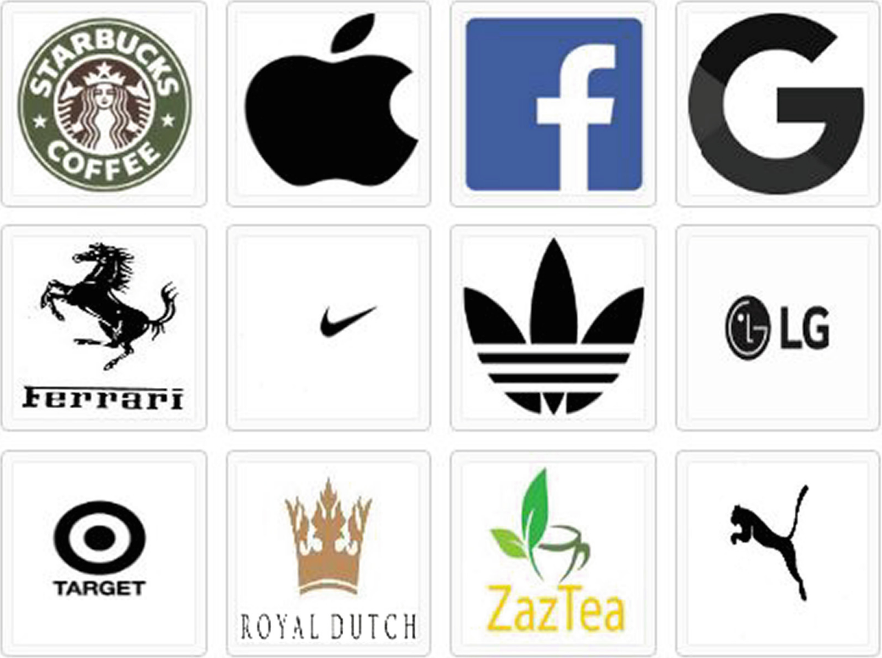

Fig. 1. Example images from USPTO trademark dataset.

2 Approaches

2.1 Content Based Image Retrieval Approach

Lucene Image Retrieval (LIRE) [21], an open source Java library was used for extracting the global and local features of the downloaded trademark images. The global features that were extracted include: Joint Color Descriptor (JCD), Pyramid Histogram of Oriented Gradients (PHOG), MPEG-7 descriptors scalable color, Color and edge directivity descriptor (CEDD) and Fuzzy color and texture histogram (FCTH). Besides this, local features were extracted based on the OpenCV implementations of SIFT and SURF. The extracted global and local image features were then stored in a Lucene index for later retrieval. For identifying similar images, LIRE either took the input query feature or extracted the feature from the input image. A linear search is then performed by reading the images from the stored Lucene index sequentially and comparing them with the input image to return a ranked order list of the best matching n candidates.

2.2 CNN Based Image Search

A total of 100,000 trademark images in Fig. 1 were downloaded from the USPTO database. The image features were then extracted by passing the images through the neural network that was pre-trained on the trademark dataset. These extracted features were used to perform image searching using the approximate nearest neighbor (ANN) variant of the nearest neighbor search (NNS) algorithm. Nearest neighbor search (NNS) is a form of proximity search that computes the distances from the query point to every single point in the target dataset and returns the data points that are closest to the query point. This technique has been successfully applied in numerous fields of applications, such as computer vision, pattern recognition, and content-based image retrieval, to name a few.

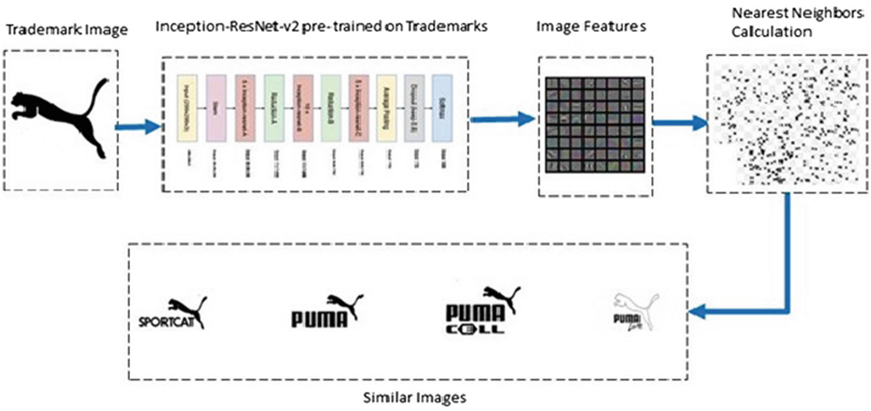
Trademark Image
Inception-ResNet-v2 pre-trained on Trademarks
Image Features
Nearest Neighbors Calculation
SPORTCAT
PUMA
PUMA
CELL
PUMA
Similar Images

Fig. 2. Image search using CNN and NNS.

For image search feature, Approximate Nearest Neighbor Oh Yeah (ANNOY) [19] and NearPy [20] libraries, were used for identifying the nearest neighbors. Each image of the trademark dataset was passed through the trained ResNet-v2 neural network as depicted in Fig. 2 to extract the intermediate representation (feature vector) of the image. These image vectors were then saved in a binary format and used to search and identify the nearest neighbors. Finally, the cosine distance between the image and the nearest neighbors was computed and then the nearest neighbors were sorted by distance to return the top K nearest neighbors.

3 Infrastructure

Amazon Web Services (AWS) cloud infrastructure and Docker [22] were utilized for performing trademark image search. AWS m4.16xlarge (64 Core Intel Xeon E5-2676 v3 Haswell processor and 256 GB DDR3 RAM) spot instance was utilized for extracting image features and then identifying the nearest neighbors of the input mark. For the machine learning approach, AWS EC2 spot instances were chosen since they have an advantage of providing surplus of computing resource at a lower price compared to the on-demand instance price. Also, Docker light weight containers were configured to ease the configuration and setup of the TensorFlow framework.

4 Results

Using a simplistic test data set (trademark variation of images from the same owner), we were able to validate the results of CNN image search approach. The test proved Mean Average Precision (MAP) score of 0.69. This sample set of comparing variations of say Puma® variations though a good starting point is not close to the complexity of the test case faced by a trademark examining attorney. We are currently in the process of obtaining a more realistic test data set curated by Trademark experts, and plan to pursue testing those (Fig. 3).

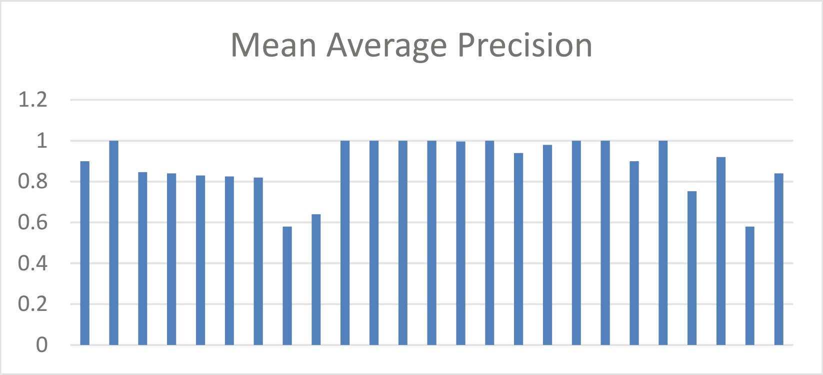

Fig. 3. MAP scores of CNN image search.

5 Conclusion

The current mechanism of searching Trademarks that depends on meta-tags such as trademark design codes, while still being the more comprehensive method of searching for likelihood of confusion in trademark images, is time consuming. By taking advantage of recent advances in Convolutional Networks (CNNs), we have been able to provide an alternate way to search for Trademarks based on image.

References

1. Patent Office Department of Commerce: Trademark Manual of Examining Procedure (TMEP). Department of Commerce, Patent and Trademark Office, Washington, D.C. (1974)
2. Bartow, A.: Likelihood of confusion. San Diego Law Rev. **41** (2004)
3. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 436–444 (2015)
4. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014)
5. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)
6. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
7. Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., Darrell, T.: DeCAF: a deep convolutional activation feature for generic visual recognition. In: International Conference in Machine Learning (2014)
8. Krizhevsky, A., Sutskever, I., Hinton, G.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, vol. 25, pp. 1106–1114 (2012)
9. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: ImageNet: a large-scale hierarchical image database. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
10. Dean, J., Corrado, G., Monga, R., Chen, K., Devin, M., Mao, M., Ranzato, M., Senior, A., Tucker, P., Yang, K., Le, Q.V., Ng, A.Y.: Large scale distributed deep networks. In: NIPS, pp. 1232–1240 (2012)
11. Szegedy, C., Liu, W., Jia, Y., Sermanet, P.: Going deeper with convolutions. In: CVPR, pp. 1–9 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR, pp. 770–778 (2016)
13. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., et al.: TensorFlow: large-scale machine learning on heterogeneous systems (2015)
14. Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I., Bergeron, A., Bouchard, N., Warde-Farley, D., Bengio, Y.: Theano: new features and speed improvements (2012)
15. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia, pp. 675–678. ACM (2014)
16. Collobert, R., Bengio, S., Mariéthoz, J.: Torch: a modular machine learning software library. Technical report IDIAP-RR 02-46, IDIAP (2002)

17. Agarwal, A., Akchurin, E., Basoglu, C., Chen, G., Cyphers, S., Droppo, J., Eversole, A., Guenter, B., Hillebrand, M., Hoens, R., Huang, X., Huang, Z., Ivanov, V., Kamenev, A., Kranen, P., Kuchaiev, O., Manousek, W., May, A., Mitra, B., Nano, O., Navarro, G., Orlov, A., Padmilac, M., Parthasarathi, H., Peng, B., Reznichenko, A., Seide, F., Seltzer, M.L., Slaney, M., Stolcke, A., Wang, Y., Wang, H., Yao, K., Yu, D., Zhang, Y., Zweig, G.: An introduction to computational networks and the computational network toolkit. Technical report MSR-TR-2014-112, August 2014. https://github.com/Microsoft/CNTK
18. Szegedy, C., et al.: Inception-v4, Inception-ResNet and the impact of residual connections on learning (2016)
19. Spotify, Annoy. https://github.com/spotify/annoy
20. NearPy. https://github.com/pixelogik/NearPy
21. Lux, M., Chatzichristofis, S.A.: LIRE: lucene image retrieval: an extensible Java CBIR library. In: ACM Multimedia (2008)
22. Merkel, D.: Docker: lightweight linux containers for consistent development and deployment. Linux J. **2014**(239), 2 (2014)