Transdisciplinary Engineering for Complex Socio-technical Systems K. Hiekata et al. (Eds.) © 2019 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/ATDE190155

# Detecting Trademark Image Infringement Using Convolutional Neural Networks

Amy J.C. TRAPPEY<sup>a,1</sup>, Charles V. TRAPPEY<sup>b</sup> and Sam C.-C. LIN<sup>a</sup>

# <sup>a</sup>Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan

<sup>b</sup>Department of Management Science, National Chiao Tung University, Hsinchu,

Taiwan

Abstract. The rapid development of consumer products with short life spans, along with fast, global e-commerce and e-marketing distribution of products and services requires greater corporate diligence to protect intangible assets such as brands which can easily be coped or placed in grey markets. Trademarks are the government registered legal intellectual property rights (IPRs) used to protect companies' brands and build brand equity. Given the rapid growth in the number of global trademark registrations, the number of trademark infringement cases is also increasing, a great challenge for the original trademark owner to detect the infringement and takes action to protect the brand image and related commercial interests. This research develops a trademark similarity assessment methodology based on the US trademark law related to the high likelihood of confusion and associated regulations. The research focuses on identifying trade mark image similarity using a deep learning approach. The convolutional neural network (CNN) and Siamese neural network (SNN) algorithms are modeled and trained using Cifar-10 and TopLogo-10 corpuses. These corpuses consist of more than 100,000 positive image pairs and more than 150,000 negative image pairs as training data. After training the model, an image input to the model extracts and recommends similar trade mark images found in the corpus. The solution assists users registering new trademarks to identifying similar marks that may lead to disputes. The solution also automatically screens images to identify marks that potentially infringe upon registered trademarks.

Keywords. Trademark Similarity Assessment, Trademark Infringement, Siamese Neural Networks, Convolutional Neural Networks

# Introduction

The protection of intellectual property involves the ownership of intangible assets such as trademarks and brand names which enhance an enterprises' global competitiveness. When consumers buy goods or services, they often rely on brand reputation and popularity. Registered trademarks are inherently valuable to ensure brands are protected as intellectual property rights. Trademarks are most valuable to original brand manufacturers since they may be bought and sold, particularly if they are well recognized by the customers. In the past few years, the numbers of global trademark applications have steadily increased. In 2017, there are more than 9 million marked registered worldwide [1]. The rapid growth in global trademarks is resulting in a fast

<sup>&</sup>lt;sup>1</sup> Corresponding Author, Email: trappey@ie.nthu.edu.tw.

increase in infringement complaints and especially with online promotions and ecommerce sales where IP (TM) rights are more difficult to monitor, trace and enforce. In order to prevent trademark infringement, increased legal action must be taken against predators which in the long run increases costs to the customers. Nonetheless, the infringement evidence is hard to quantify especially if multiple Internet sites are used for illegal usages of the marks. The most serious punishment for trademark infringement cases may lead to short term imprisonment, which is a minor penalty considering the huge amounts of damages caused to the brand equity [2]. This research intends to develop and implement an intelligent trademark similarity assessment method and system based on criteria of US law's definition of brand or trademark confusion [3].

## 1. Trademark infringement

A trademark is a sign or label used to represent the identity of given products or services. Trademarks may be expressed as words, images, colors or even sounds and scents. When an enterprise or individual successfully registers a trademark in a country's intellectual property registering agency, the owner will have the exclusive right to use the mark for a given period of years with rights to renew. Trademarks are a type of intangible asset which must be properly managed to create profits. The goal is to use a trademark to strengthen the customers' brand loyalty which in turn increases sales, creates premium price and profit margins that help the customer to distinguish between competing brands. Trademark infringement means that the infringer uses the same or similar names or images on similar goods and service offerings without permission or a licensing agreement from the trademark owner. The Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) [4]; a multilateral treaty under the World Trade Organization (WTO), defines intellectual property rights regulations. Article 16 of TRIPS states that trade mark registrants shall have exclusive rights to prevent others from using the same or similar marks on similar goods or services. This behavior is called "likelihood of confusion."

In addition to TRIPS, international trademark agreements are also implemented by World Intellectual Property Organization (WIPO) through the Paris Convention and Madrid Agreement. These treaties and agreements have clearly defined the definition and norms of trade mark infringement. The Paris Convention was established in late 1800 and has been revised seven times [5]. The Paris Convention is the first international convention on intellectual property rights and the Madrid Agreement was established in 1891 [6]. The agreement facilitates the registration of trademarks in different member countries around the world.

#### 2. Artificial intelligence and neural networks

The term "Artificial Intelligence" (AI) was first proposed in the 1950s [7]. The definition of AI is often referred to the ability of machines to understand, think, and learn in a similar fashion as human beings. During the 1970s, AI began to lose popularity since it failed to achieve the expectations that were anticipated to mimic or duplicate human intelligence. AI machines at the time had insufficient memories and computing power to acquire large amounts of knowledge and solve complicated text

processing problems. After the 1980s, AI has been extended to other research fields, including machine translation, expert systems, game theory, pattern recognition, machine learning, and intelligent control. AI methodologies are subdivided into strong AI and weak AI. A strong AI machine is considered to have human-like intelligence and performs in a manner similar to a human. Weak AI refers to a machine capable of performing specific tasks and achieving results as well as humans. After the year 2000 the growth of the Internet, the prevalence of sensors, the emergence of big data, the development of e-commerce, the interconnection and integration of data and knowledge in social, physical and cyberspace provided researchers with a better environment to solve AI problems.

AI has many different applications and each application may require specific AI techniques and solutions. Among them, neural network modeling and machine learning approaches have been very successful. A standard neural network consists of many simple, connected processors called neurons, each producing a sequence of real-valued activations. Schmidhuber [8] collated and reported the history of neural network development. Around 1960, the first single-layer neural network was proposed, called the perceptron. Then, neural network modeling began to flourish and many improvements followed. During the 1980s, recurrent neural networks and convolutional neural networks were introduced for solving specific problems. After the year 2000, with the advancement of more efficient and less expensive processing units, deep neural networks were applied to speech recognition, computer vision, financial forecasting, and pattern recognition.

## 3. Image-based similarity algorithms

There are many ways to measure image similarity. First, there is a need to define image features. Features are visual aspects (such as a star shape) of an image that have similarity and can be detected (identified) and measured between images. Image similarity is used to score the degree of similarity between the two images. This section discusses the image extraction method and is divided into two subsections. The traditional feature extraction method, called manual feature extraction, is explained in Section 3.1. Deep learning feature extraction is discussed in Section 3.2.

## 3.1. Manual feature extraction

The most common manual image feature extraction methods are divided into color, texture and shape. Color is one of the most prominent features of an image and is an important visual measure that is easily measured as different or similar by humans. Researchers like to use color since it is unaffected by the natural rotation, scaling and translation of the image. Many methods can be used to describe color features, such as color histograms and a color coherence vector [9]. Color histograms quantify the color within the image and calculate the number of pixels for each color, and then find the mean and standard deviation from the histogram. The color coherence vector is a color feature that contains spatial information about the color distribution. This method has the advantage of expressing the spatial positions of the image colors which is not provided by the color histogram. Color moments are an effective feature used to differentiate images. The measurements most frequently used to build color feature vectors include mean, variance, and skewness [10].

For texture-based feature extraction, the local binary pattern is used for texture classification. This method compares each pixel with neighboring pixels and saves the result as a binary digit [11]. The Gabor Wavelet texture is noted for its effectiveness in extracting image textural features. The method analyzes whether the image has specific frequency content in a specific direction of a specific region [12]. The intensity of the wave represents the pixel, and the frequency and direction of the wave represents the position of the pixel. For shape-based features, Hu moments and eccentricity are commonly used [10]. The Hu moment considers the gray value of the image as a twodimensional or three-dimensional density distribution function. The Hu moment contains seven invariant moments, and these invariant moments utilize image translation, expansion and rotation invariance. Eccentricity is an important way to describe the shape of an area by calculating the mass or edge points of the area. The Zernike moment is similar to the Hu moment and describes the shape of an image by using an orthogonal representation of two lines in a right-angle [13]. SIFT is a computer vision algorithm used to detect and describe the local features of an image. The method is based on interest points and their local appearance on the object and is unrelated of the size and rotation of the image.

## 3.2. Deep learning feature extraction

The emergence of neural network modeling and deep learning techniques has created a wide range of intelligent applications for advanced feature extraction. In the field of image recognition, deep learning technology plays a dominant role. Among image recognition solutions, convolutional neural networks (CNN) significantly increase accuracy in image recognition and often outperform humans given similar tasks. The first generation of CNN was called Neocognitron [14]. The CNN model is a deep learning model based on the human brain visual organization. The algorithm uses low-level features of the underlying image layer, and then slowly combines the lower-level features toward the learning of higher-order features. CNN typically consists of three network layers. The first is the convolution layer, the second is the subsampling layer (or pooling layer), and the third layer spreads the pooling layer into a one-dimensional form called fully connected layer.

Using CNN for image similarity comparison is an important application. There are numerous research papers in the area of CNN image similarity. Liu et al. [15] propose deep relative distance learning which uses a bifurcated deep CNN to project the original vehicle image into Euclidean space to measure the similarity of any two automobiles. Some researchers have proposed using the Siamese Neural Network (SNN) combined with CNN architecture to measure the similarity between images. For example, Bell and Bala [16] use the combined SNN and CNN architecture to explore product similarity using image embedding. Their research demonstrates several applications of visual search, such as finding products with similar shapes. In addition, Melekhov et al. [17] also apply SNN architecture to mark images in image database to achieve better performance in similar images matching.

Wang et al. [18] propose a depth ranking system to measure the similarity of fine images. The objective is to use a triplet-based network model architecture for image similarity comparison. Our research paper uses a multi-scale CNN structure to improve the accuracy of image feature extraction. Multi-scale CNN is often used for image similarity comparison. The advantage is that CNN jointly learns image embedding from top and bottom layers. For the task of image similarity, the multi-scale model is demonstrated to be an improvement over traditional CNN [19].

#### 4. Research methodology and structure

This study uses the integrated Siamese Neural Network and CNN to construct the trademark similarity assessment model. The feature vector generated by the CNN approach is used for comparing whether two images are similar. The CNN model uses the SNN architecture to learn the similarity of images from matched and unmatched image pairs [17] [20]. The research architectural flow is shown in Figure 1. The structure consists of two identical branches that share weights and parameters and are connected at one or more layers. The SNN receives case pairs as input for the training and testing phases to develop knowledge at the object-to-object level. The main goal of the SNN structure is to transform the input image into a feature vector ( $F(I_1)$ ,  $F(I_2)$ ), and then minimize the vector distance of the matched image pairs and maximize the distance of the unmatched image pairs. The construction process of Siamese Neural Network includes developing a training data set, CNN structure design, and adjusting the loss function and training parameter settings.



Figure 1. The architectural flow of key research steps.

#### 4.1. Training data collection

The first step to build the SNN architecture is collect suitable images and creating a data archive. The SNN model requires pairs of positive images and negative images. A positive image sample pair represents the same or similar types of figures and a negative sample pair is the opposite. A large number of image pairs and labels from the Cifar-10 and TopLogo-10 image datasets are used as the training datasets [21] [22]. The Cifar-10 dataset covers a total of 60,000 32x32 color images and is classified into 10 classes, each containing approximately 6,000 images. The TopLogo-10 dataset includes 10 unique logo categories related to the most popular brands of clothing, shoes and accessories. There are 70 images in each logo category and includes a total of 700 logo images. This study uses data augmentation techniques to improve the training data set. For each image, we generate a duplicate image that is shifted, zoomed in/out, rotated and flipped. All images located in the same class are used to form positive label image pairs, while images located in different classes are used to form negative label image pairs. The study collected more than 250,000 image pairs, containing 100,000 positive image pairs and 150,000 negative image pairs as CNN training data.

# 4.2. CNN architecture design

The SNN of this study covers two CNN branch structures. The network architecture is based on the VGGnet architecture from the Visual Geometry Group (VGG) at Oxford University [23]. The main contribution of this architecture is to enable deep learning for the CNN's through smaller convolution (3x3) stacking and smaller pooling (2x2) compression to enhance the accuracy of results. In addition, the VGGnet architecture effectively captures complex and expressive image features by using smaller convolution kernels. In our research, we also improve the CNN architecture incorporating multi-scale and random filtering concepts to optimize the model outcome.

# 4.3. Loss function establishment

After passing through the CNN network, the feature vectors obtained by the two branches are output to the loss layer. The contrastive loss function model attempts to minimize the distance between the features of the positive image pairs and maximize the distance between the negative image pairs. The contrastive loss function is shown in Equation (1).

$$L = \frac{1}{2N} \sum_{n=1}^{N} y D^2 + (1 - y) max (Margin - D, 0)^2$$
(1)

where N is the number of samples and D is the Euclidean distance between the two feature vectors generated by the CNN. The mathematical expression can be expressed as  $D = || F (I_1) - F (I_2) ||^2$  where y = 1 indicates that the image pair of the two feature vectors belong to the positive label, and y = 0 indicates that the image pair of the two feature vectors belong to the negative label. The Margin in Equation (1) is the margin between the positive and negative image pairs and the value is determined empirically. Margin is similar to a threshold value indicating that the loss function only considers the case where the Euclidean distance of the negative image pairs is between 0 and the Margin. When the distance exceeds the Margin, the loss value is assigned 0. A larger margin pushes positive and negative image pairs further apart. In our study, *the* margin is set as 1. According to the general design rules of the contrastive loss function, the matching pairs must be closer together in the feature space while pulling the distance of the non-matching pairs further apart.

# 4.4. Training parameter setting

The CNN/SNN models are trained on a GPU server using an open deep learning framework. The framework consists of the *Keras* neural network API that runs at the back end of TensorFlow, CNTK and Theano. This study uses keras.models, keras.layers.cores, keras.layers, and keras.models modules to build the core architectural elements of the CNN model. Among these modules, there are many important packages. For example, the Sequential function passes a layer list to construct the model. Density measures the number of fully connected layer neurons and Activation is the activation function setting. The Dropout rate measures neurons discarded when updating parameters during training to prevent overfitting. Flattening is used to flatten the input and convert the multidimensional input into a one-dimensional formation. Convolution2D is the function used to build the convolution layer and MaxPooling2D is the function used to build the pooling layer. The python code, developed in this research, can be provided upon request. In addition to the above

parameters, the learning rate, optimizer, batch size, and epochs are additional and important training parameters. The learning rate is the step size used during forward and backward propagation during network training. The optimizer speeds up network training and this study uses the Adam optimizer [24]. The batch size is the sample size used for 1 iteration and epochs measure the number of times all training data complete the parameter setting operations. Table 1 shows the value of important hyperparameters in the neural network model construction of this study.

Hyper-parameter	Value	
Epochs	40	
Batch size	256	
Dropout	0.5	
Optimizer	Adam	

Table 1. The hyper-parameters of the neural model construction.

#### 4.5. Experimental results for the CNN/SNN models

In the process of training the model, the study divides the image data corpus into a training set and a validation set. The training data consist of 250,000 image pairs and 20,000 image pairs are used as the validation data. The experiments are run using different model architectures. By observing the loss value and accuracy value of both the training and validation data, the best performing model architecture is selected. Table 2 shows the experimental results. When the similarity is greater than the predefined threshold (0.6), the image is categorized into the similar image pair.

Table 2. Experimental results of models.

Model	Training set	Validation set
	accuracy	accuracy
CNN (VGG top 3 layer) - baseline	71 %	79 %
CNN (VGG top 7 layer)	76 %	85 %
CNN (VGG top 9 layer)	76 %	84 %
Multi-scale CNN (3 layer + 5 layer)	76 %	89 %
Multi-scale CNN (3 layer + 7 layer)	76 %	89 %
Multi-scale CNN (3 layer + 9 layer)	73 %	87 %
Multi-scale CNN with random filters (3 layer + 9 layer)	73 %	86 %

#### 5. Case study and result

The study uses fifty trademark infringement cases to evaluate the effectiveness and accuracy of the proposed CNN/SNN solution. By comparing the system results to the legal judgments determined by the courts, we evaluate the effectiveness of the approach. The results of the court rulings are divided into three types. The first is that the trademark image between the plaintiff and the defendant constitutes infringement, the second is that the legal judgment shows no infringement, and the third outcome is an appeal with an undecided and ongoing legal process. Among the 50 trademark disputes relating to image similarities and confusion, 40 case results are infringement and the other 10 are no infringement.

First, we discuss 10 cases in which the infringement is not established. As shown in Table 3, the last column indicates the similarity measurement using the proposed CNN/SNN method. The experiment results found that, when investigating the evidence of infringement, the court will explore other factors in addition to the image similarity

between trademarks. The additional evidence includes strength and application of the disputed marks, the relatedness of the goods, defendant's use of the mark, and evidence of actual confusion. For example, Cases 1, 6, 7, 10, the judgments indicate that there is insufficient evidence of actual confusion, yet the image similarity measures are about or above the threshold level (0.6). In Cases 2 and 3, the judge considered that the plaintiffs' trademarks are weak in strength and recognition. Thus, there was insufficient evidence to support the violation of uniqueness. In Case 8, the judge considered that the scope of services for both trademarks were different, so infringement was not established. In Case 9, the judge considered that the defendant's environmental protection trademark was unique with its own merit, so the plaintiff's complaint was not accepted.

	Plaintiff	TM Image	Defendant	TM Image	Similarity
Case 1	Cosmetic Derms. & Vein Cent.	{	New Faces Skin Care		0.628
Case 2	Starbucks	APR BUCCO	Starpreya		0.785
Case 3	Edison S.p.A.	9	Eolus Vind AB	9	0.502
Case 4	Crocodile International Pte Limited	Crocodile	Lacoste		0.703
Case 5	Sieun Ha	X	Louis Vuitton Malletier	X	0.766
Case 6	Shape Shopfitters Pty Ltd	SHAPE	Shape Australia Pty Ltd	SHAPE	0.595
Case 7	Guangzhou Nandadi Textile Garment Co., Ltd	25	RCRV, Inc	and the second	0.808
Case 8	Apple	Ś	Apfelkind	Ő	0.526
Case 9	Apple	Ú	GreeNYC	Ň	0.659
Case 10	Sony Ericson	6	Clearwire Corporation	C	0.704

Table 3. Trademark infringement cases with no infringement determined by court.

Regarding the 40 trademark infringement cases, the similarity assessment results are shown in Table 4. The threshold values of similarity are divided into 0.5 and 0.6 for comparison. In the model architecture, it is divided into single-scale CNN, multi-scale CNN and multi-scale CNN with random filters. From the results, single-scale CNN has no significant change in accuracy regardless of the number of network layers. The

multi-scale CNN is better at finding fine-grained image similarities than traditional CNN's, especially with regards to the trademark images. Therefore, multi-scale CNN achieves better image expression ability. Neural networks with random weights refer to the random selection of the weight between the hidden layer and the input layer. This study finds that applying random weights to the model architecture lowers training complexity in comparison to traditional feed-forward neural networks.

Model	Accuracy / (40 cases)	Accuracy / (40 cases)
	Similarity threshold: 0.5	Similarity threshold: 0.6
CNN (VGG top 3 layer) - baseline	38/40	30/40
CNN (VGG top 7 layer)	38/40	30/40
CNN (VGG top 9 layer)	36/40	30/40
Multi-scale CNN (3 layer + 5 layer)	38/40	33/40
Multi-scale CNN (3 layer + 7 layer)	38/40	33/40
Multi-scale CNN (3 layer + 9 layer)	39/40	36/40
Multi-scale CNN with random filters (3 layer + 9 layer)	40/40	35/40

Table 4.	Test results	of TM	infringement	cases.

# 6. Conclusion

This study uses the deep learning method to perform the automatic image similarity assessment. The methodology is applied as an e-discovery tool for gathering evidences for trademark infringement litigations, considering disputed TM image similarities. This study uses deep learning CNN/SNN modeling architectures to perform image similarity assessment applied to trademark image infringement case analysis. SNN is used to construct the similarity model using training data of regular images and trademark images. Based on VGGnet, the CNN architecture is designed and the effects of different network layers on the results are varied and compared. The results show that multi-scale CNN has an improved similarity comparison with better image embedding learning results. For the 50 trademark infringement cases, the results of the final judgments are compared with the methods proposed in this study. The image similarity assessment solution yields great insights in identify infringement based on image confusion. The research also demonstrates that trademark infringement legal judgment often considers evidence directly related to the market environment in dispute, such as the strength and applications of the marks and the obvious intention of trademark infringement, in combination with the image similarity measure. So far, 250,000 training sets may not be sufficient to train CNN/SNN models. We will continue to expand training data and improve the model architecture in the future research. In addition, further research is also being pursued to consider multiple aspects of trademark infringements, e.g., spellings and pronunciations of marks.

## Acknowledgement

This research is partially supported by the Ministry of Science and Technology research funding (MOST-107-2221-E-007-071 and MOST-107-2410-H-009-023) in Taiwan.

# References

[1] World Intellectual Property Organization (WIPO), 2018, World Intellectual Property Indicators 2018,

Accessed: 15.10.2018. [Online]. Available:

- http://www.wipo.int/edocs/pubdocs/en/wipo\_pub\_941\_2017.pdf
- [2] TIPO, 2018, Agreement on trade-related aspects of intellectual property rights, Accessed: 15.10.2018. [Online]. Available: https://www.tipo.gov.tw/ct.asp?xItem=207100&ctNode=6780&mp=1
- [3] Lanham Act, 2016, Lanham (trademark) act index, Accessed: 20.11.2018. [Online] https://www.bitlaw.com/source/15usc/index.html.
- [4] World Trade Organization, 2017, WTO A Handbook on the WTO TRIPS Agreement (Agreement on Trade-Related Aspects of Intellectual Property Rights), Accessed: 8.2.2019. [Online]. Available: https://www.wto.org/english/res e/publications e/handbook wtotripsag12 e.htm.
- [5] United International Bureau for the Protection of Intellectual Property (BIRPI), 1968, Paris Convention for the protection of industrial property, Accessed: 13.02.2019. [Online] https://www.wipo.int/edocs/lexdocs/treaties/en/paris/trt\_paris\_001en.pdf.
- [6] WIPO, 2016, *The Madrid system for the international registration of marks*, Accessed: 13.02.2019. [Online] https://www.wipo.int/publications/en/details.jsp?id=4045
- [7] Y. Pan, Heading toward artificial intelligence 2.0, Engineering, vol. 2, no. 4, 2016, pp. 409-413.
- [8] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural networks*, vol. 61, 2015, pp. 85-117.
- [9] K. Roy, and J. Mukherjee, Image similarity measure using color histogram, color coherence vector, and sobel method, *International Journal of Science and Research (IJSR)*, vol. 2, no. 1, 2013, pp. 538-543.
- [10] F. You, and C. Zhang, The technique of color and shape-based multi-feature combination of trademark image retrieval, in *Proceedings of the International Conference on Multimedia Technology*, Ningbo, China, 2010, pp. 1-5.
- [11] N. Gupta, & V. A. Athavale, Comparative study of different low level feature extraction techniques for content based image retrieval, *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol. 1, no. 1, 2011, pp. 39-42.
- [12] N. Gupta, V. A, Athavale and M. I. Khan, Graphical user interface of efficient image quality assessment using new similarity metrics, *Digital Image Processing*, vol. 2, no, 2010, pp. 342-347.
- [13] F. M. Anuar, R. Setchi, Y. K. Lai, Trademark image retrieval using an integrated shape descriptor, *Expert Systems with Applications*, vol. 40, no. 1, 2013, pp. 105-121.
- [14] K. Fukushima & S. Miyake, Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, *Pattern recognition*, vol. 15, no. 6, 1982, pp. 455-469.
- [15] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, Deep relative distance learning: tell the difference between similar vehicles, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, Nevada, USA, 2016, pp. 2167-2175
- [16] S. Bell and K. Bela, 2015, Learning visual similarity for product design with convolutional neural networks, Accessed: 15.10.2018. [Online]. Available: https://www.cs.cornell.edu/~kb/publications/SIG15ProductNet.pdf
- [17] I. Melekhov, J. Kannala, and E. Rahtu, Siamese network features for image matching, Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 2016, pp. 13-18.
- [18] J. Wang, Y. Song, T. Leung, C. Rosenberg, J. Wang, J. Philbin, Y. Wu, Learning fine-grained image similarity with deep ranking, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, 2014, pp. 1386-1393.
- [19] S. Appalaraju, and V. Chaoji, 2017, Image similarity using deep CNN and curriculum learning, Accessed: 15.10.2018. [Online]. Available: https://arxiv.org/abs/1709.08761
- [20] S. Zagoruyko, and N. Komodakis, Learning to compare image patches via convolutional neural networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), Boston, MA, 2015, pp. 4353-4361.
- [21] A. Krizhevsky, and G. Hinton, 2009, *Learning multiple layers of features from tiny image*, Accessed: 15.10.2018. [Online]. Available:
- http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.222.9220&rep=rep1&type=pdf [22] H. Su, X. Zhu and S. Gong, Deep learning logo detection with data expansion by synthesising context,
- 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, pp. 530-539.
  [23] K. Simonyan, and A. Zisserman, 2014, Very deep convolutional networks for large-scale image recognition, Accessed: 15.10.2018. [Online]. Available: https://arxiv.org/abs/1409.1556.
- [24] D. P. Kingma and J. Ba, 2014, Adam: A method for stochastic optimization, Accessed: 8.1.2019. [Online]. Available: https://arxiv.org/abs/1412.6980