Robust AI Digital Copyright Protection Scheme

Balika J Cheliah, Srushti Bompelli, Madhumita Sampath, Maithili Ghogre, Aravind Ajithkumar

Abstract: The way we watch and consume media is constantly changing. With growing Web Technologies, creation, marketing and distribution of media content is fast changing. This opens up immense opportunities for piracy and redistribution. One can find several ways to share media content online like Social Networking Portals, Free Cloud Spaces and Drives, emails, chats etc. Due to the sheer volume of information and content available on the Internet, it becomes impossible to detect and stop piracy manually. Thus, the objective of this paper is to propose a solution to fight piracy using Artificial Intelligence and Machine Learning. The proposed system leverages Artificial Intelligence and Machine Learning by using content monitoring solutions that search and identify pirated content on the internet. This is done by identifying the original source of distributed content based on the visual information present in the image such as the broadcaster logo. The paper covers practical issues around building a system with its workflow, training and performance.

Keywords: Artificial Intelligence, deep learning, media, logo detection

I. INTRODUCTION

With the advent of the Internet, media and content is shared online now, more than ever. The Internet has made it easy for sharing content amongst each other. Content that we used to pay for to watch on TV is now available illegally on the Internet. Sports matches, movies, shows, series, documentaries, etc. are illegally broadcasted on the Internet. Illegal redistribution of content, thus poses a serious threat to the content production business today. Through this paper, we present a solution that can help fight piracy with the help of certain AI and Machine Learning Techniques.

The affordable pirate technologies such as Periscope and Meerkat, combined with various streaming devices have made it possible to redistribute content online. For example, consider the highly grossed show Game of Thrones telecasted on HBO. HBO telecasts only one episode per week. The episode has value only if it's watched when it is telecasted. Capitalizing on this, many pirates with legitimate subscriptions to the channel, record or stream the episode online via various streaming protocols that are available.

Another common example of such piracy is the illegal rebroadcast of sports matches. The sports matches have a high value when they are live. Making use of this, pirates stream the content online either via web streaming or they upload to file sharing networks such as bittorrents etc. The content undergoes limited modification such as distortion of the logos, changing meta data, encrypting and compressing the content so that it is difficult to find out whether the broadcast is legitimate or not. But such forms of piracy are extremely hard to detect. One cannot find out which streams are decoys and which streams are pirated.

Metadata provided along with the streams can be used to find out whether it is the original content or not. But the metadata is often unstructured and inconsistent. This means that one has to analyze the entire stream to find the original source. When analyzing the stream, the logo of the broadcaster is the only indication of the original source of the stream. Once the logo has been identified, the stream can be traced back to its original source and then the pirated content can be removed by taking certain measures. Detection and recognition of logos has been made harder by distorting them, reducing their quality, hiding them intentionally or by applying multiple layers of encoding.

Through this paper, we are trying to provide an approach using Machine Learning and Artificial Intelligence that will detect and recognize broadcaster logos from video streams. We explain the workflow of such a system along with the necessary architectures that we have selected and the reasons for choosing them over the existing architectures. The system we propose is customized for logo detection and recognition and we hence compare its accuracy and performance with the existing systems.

II. RELATED WORK

A lot of research has been done in the field of media piracy. The paper on "Analysis of Watermarking Techniques in Video" suggests using certain watermarking techniques which will distinguish the original video from the duplicate.

Revised Manuscript Received on April 14, 2019.

Balika J Cheliah, Department of Computer Science and Engineering SRM Ramapuram, Chennai, India

Srushti Bompelli, Department of Computer Science and Engineering SRM Ramapuram, Chennai, India

Madhumita Sampath, Department of Computer Science and Engineering SRM Ramapuram, Chennai, India

Maithili Ghogre, Department of Computer Science and Engineering SRM Ramapuram, Chennai, India

Aravind Ajithkumar, Department of Computer Science and Engineering SRM Ramapuram, Chennai, India

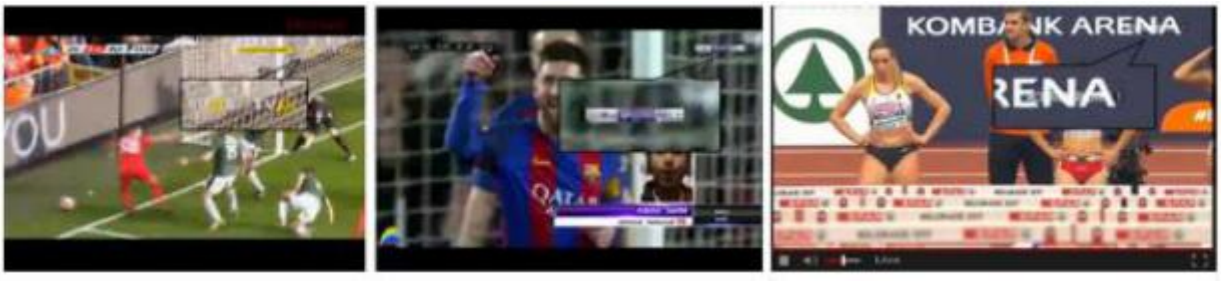


Fig. 1 Distorted broadcaster logos from pirated online streams

The watermark symbols can be embedded within the videos which will later enable us to detect copyright. But many of the existing algorithms focus on robustness and intangibility instead of making the algorithm useful for real time applications.

Logo detection and recognition have found multiple uses in recent times. Various research has been conducted on logo detection of images for classification, detection of copyright infringement, marketing and advertisements, for intelligent traffic control systems etc. According to the paper, SSD: Single Shot Multibox Detector, SSD is a method for detecting objects in images using a single deep neural network. As the name suggests, it uses only a single shot to detect multiple objects in an image which is therefore faster than conventional CNNs.

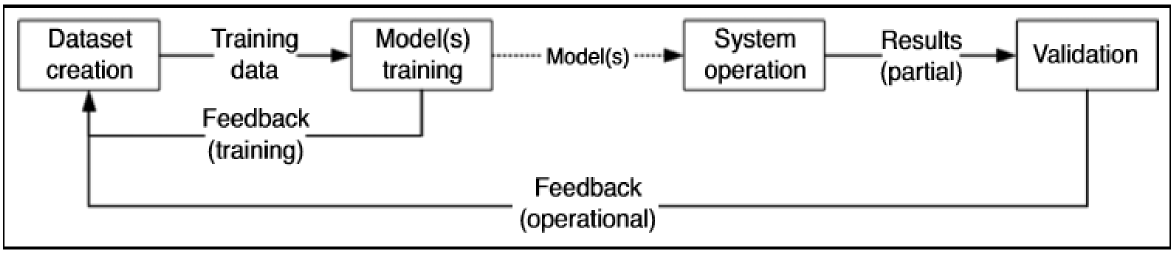


Fig 2. Workflow for a machine learning system

Zhe Li in their paper titled, “Fast Logo Detection and Recognition in Document images suggested a method which allows segmentation-free and layout-independent logo detection and recognition. The method was tested on several real time documents that showed an increase in time and performance. But their algorithm was restricted only to documents and its usage for logo detection was not possible.

Recent developments have begun using Convolution Neural Networks for Image recognition tasks. CNNs are a subset of Neural Networks. They are currently the best available resource for image extraction and recognition. They use modern machine learning and computer vision methods to extract information from images. But one major issue with CNNs is the training. They require a large data set that contains all the possible outcomes. After much research into the various techniques available, we settled upon

CNNs as they seemed to be able to perform genetic feature extraction when processing images. The same architecture can be used for multiple visual recognition tasks just by modifying the training set. Using the knowledge gathered about CNNs and logo detection from various resources we build a system that would recognize logos and images from video streams.

III. PROPOSED FRAMEWORK

We propose a system that can visually detect and recognize logos. The input to the system is a frame from the broadcast stream. The system detects whether the frame contains a logo and then identifies the broadcaster for that logo. The system is expected to work in noisy environments, that is, the system is expected to recognize distorted, partially hidden, compressed images with low image quality. This system is implemented with machine learning and CNNs.

The existing systems cannot be used because none of the systems were trained with the logos we require nor were they trained to detect logos in distorted environment. Thus, data set creation becomes an important task. The second most important part of the system is architecture selection. There are multiple CNN architectures available that can be used for image detection and recognition but the one that serves our purpose best must be identified and selected. We therefore use a custom architecture called as Logonet.

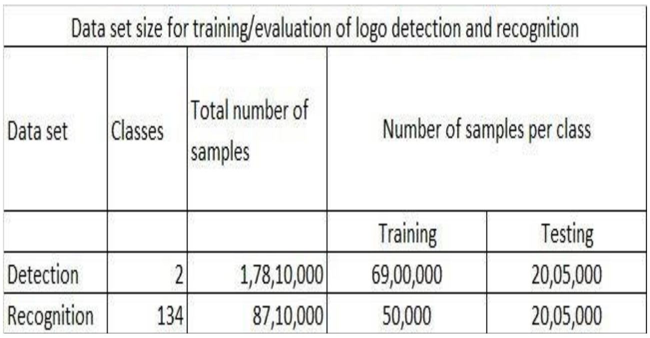
A. Data Set Creation

There is a large amount of video content online from which we can create our datatset. There are no pre existing datasets that can directly be used to train our model. Hence, we make use of the online resources to obtain our dataset. We begin by finding relevant videos or streams online. The next step is to capture frames that contain the logo of the broadcaster. After the required frames are captured, the logos have to be extracted. For this purpose, we mark the area around the logo and then extract it. To increase the size of the data set, the extracted images undergo data augmentation. This is done to shift the images in various directions so that the logo does not always appear at the center of the frame.

Scaling is also done to produce zoom effect which increases the reliability of the dataset. Also, synthetic images are built by inserting logos in samples that did not originally contain logos.

The dataset is arranged into two collections. The first one will be used for Logo detection whereas the second one will be used for logo recognition. The LogoDetect dataset will contain two classes, one with samples of known logos and the other will be a noise class, that is, it will contain samples without the logos. Similarly, for logo recognition, there are samples of each of the broadcaster logos as well as a supplementary noise class. Addition of a noise class to a dataset helps make the training process robust and improves performance and efficiency of the models.

Table. 1 Dataset for training the logo detection/ recognition models



Data set size for training/evaluation of logo detection and recognition

Data set	Classes	Total number of samples	Number of samples per class	
			Training	Testing
Detection	2	1,78,10,000	69,00,000	20,05,000
Recognition	134	87,10,000	50,000	20,05,000

B. Architecture Selection

CNNs are the desired method for detection and recognition, given their flexibility and ease of usage. There aren't any conclusive results about the efficiency and real time optimality of the various CNN architectures that are present. Any generic CNN architecture can be used for image recognition as there isn't a custom architecture for logo detection and recognition.

But this poses a few problems. Any generic architecture comes with a pre trained model that can be easily retrained for the given task. But there is no assurance that the selected architecture would be optimal for the task. Another problem is that these pretrained models require the input image to be224x224 pixel but this isn't suitable for our task. The last point of concern is that such pretrained models may contain some amount of bias which an attacker might make use of by arbitrarily crafting images that could result in inaccurate recognitions.

Thus, we propose to use a custom architecture. A custom architecture designed and trained particularly for logo detection and recognition would be ideal for the task. It would also require smaller amount of resources to train and test as opposed to a number of days for the existing architectures. But designing a CNN architecture requires extensive research and investigation.

Logonet is our proposed architecture. The input size, number, type of layers has been modified to fit the task at hand which is logo detection and recognition. The Logonet consists of the following components:

- 1 Convolution with 11 × 11 kernel size (1CONV)
- Rectified Linear Unit Layer Activation (RELU)
- Response Normalization Layer
- 1 Maximum Pooling (4 × 4 kernel)
- 2 Convolution with 5 × 5 kernel size (2CONV)
- Rectified Linear Unit Layer (RELU)
- Response Normalization Layer
- 2 Maximum Pooling (3 × 3)
- 3 Convolution with 3 × 3 kernel size (3CONV)
- Rectified Linear Unit Layer Activation (RELU)
- 4 Convolution with 3 × 3 kernel size
- Rectified Linear Unit Layer Activation (RELU)
- 3 Maximum Pooling (3 × 3)
- Fully Connected Layer (4096 nodes)
- Rectified Linear Unit Layer Activation (RELU)
- Fully Connected Layer (4096 nodes)
- Rectified Linear Unit Layer (RELU)
- Soft-max out

The implementation is a three-step process. First, a module called proposal generator segments the image because CNN does not allow images with high resolutions. Second, these segmented images are then passed through the logo detection model which identifies whether these segments contain the logos or not. Third, after detection it is passed to the recognition model which classifies the candidate segments. The architecture of logo detection and recognition is shown below.

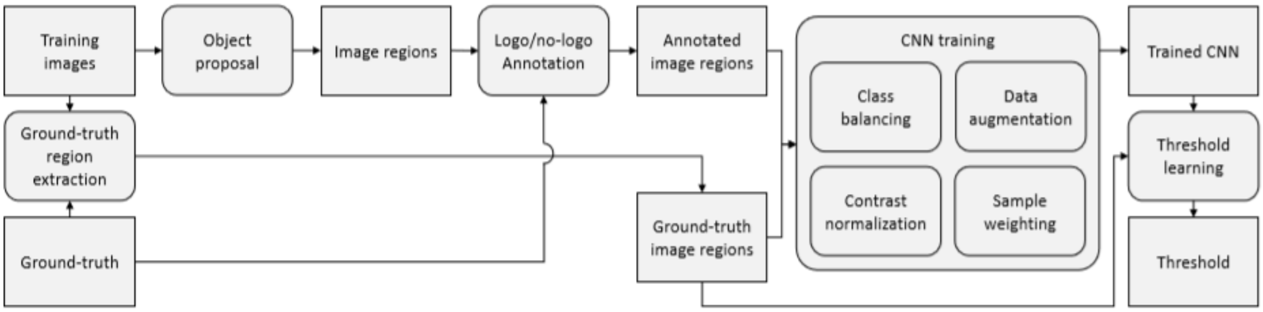


Fig. 3 Training framework for logo recognition

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

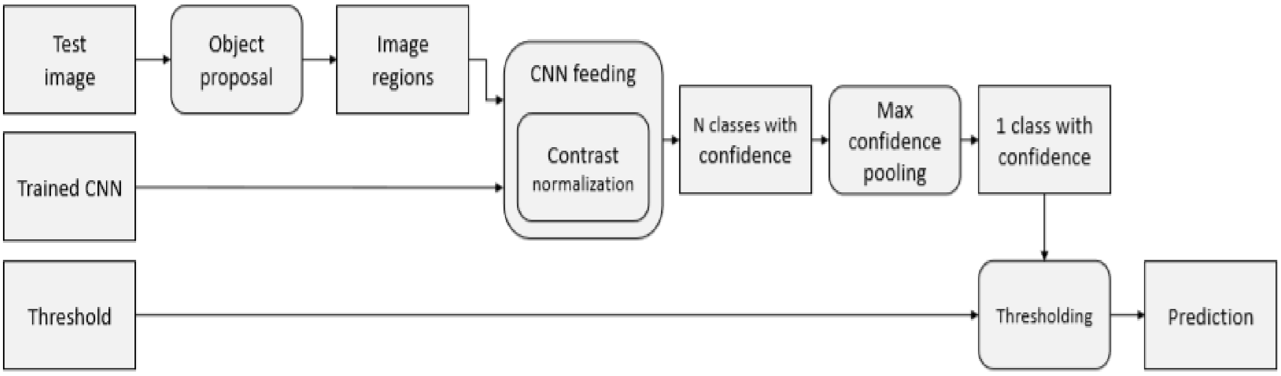


Fig. 4 Testing framework for logo recognition

IV. TESTING AND RESULTS

The custom architecture Logonet is used to train and classify the dataset. The dataset is also trained with existing architecture, AlexNet and the comparison is recorded in Table 2.

Objective metrics was utilized to evaluate the performance of this system. Precision, Recall and F1 Score are the metrics that we have used to measure the performance of the framework. These metrics are computed in terms of correct/incorrect recognition of the broadcaster logos for each class. At the end, the results are averaged to obtain the final value.

Precision (P) represents the ratio of correctly identified logos to the total number of logos detected. It can be represented as

$$P = \frac{TP}{TP+FP}$$

P is True Positive and FP is False Positive. TP refers to correctly recognized logos where FP refers to the logos that have been misclassified.

Recall (R) represents the ratio of true detection of the logos over a total number of logos in the video and computed as:

$$R = \frac{TP}{TP+FN}$$

F1-score represents the harmonic mean of precision and recall. This is a very useful metric for performance comparison in cases where some methods have better precision but lower recall rate than the other method. In this scenario, precision and recall rates independently are unable to provide true comparison. Therefore, F1-score can reliably be used in such cases for performance comparison. F1-score is computed as:

$$F1 = \frac{P*R}{P+R}$$

Performance comparison of the proposed method with the standard classifiers such as AlexNet and LogoNet shows that the proposed method has an average F1 score of 98.41%. The score is marginally lesser than the state-of-the-art architectures.

Table. 2 Comparison of accuracy with existing architecture and custom architecture

Task	Network		Training			Testing	
		Precision (%)	Recall (%)	F1 Score (%)	Precision (%)	Recall (%)	F1 Score (%)
Detection	**AlexNet**	99.508	99.508	99.508	99.078	99.072	99.072
	Logonet	99.213	99.213	99.213	98.817	98.817	98.817
Recognition	**AlexNet**	99.585	99.585	99.585	99.48	99.48	99.48
	ResNet	99.646	99.645	99.645	99.577	99.575	99.576
	Logonet	98.544	98.512	99.523	99.438	98.407	98.418

This shows that the custom architecture is highly reliable and accurate and with few modifications and improvements can be used for logo detection and recognition and it would be expected to produce reliable output.

V. CONCLUSION

Thus, a robust model has been proposed to protect digital content online. Using the custom architecture Logonet, pirated content can be detected. The dataset for this model should be increased substantially to include all the broadcasters present. Research has to be made about copyright laws in various states and whether rebroadcasting is permitted or not. Future works for this proposed framework would include automatically detecting the stream and if pirated content is found then the respective internet service provider should be contacted and the video should be removed. Artificial Intelligence and Machine Learning are growing fields. Extensive research must be conducted in the fields of deep learning to help fight piracy which is a growing threat to the content production and distribution business.

REFERENCES

1. D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” Proc. Conf. Comput. Vision, Vol. 2, pp. 1150–1157,1999.
2. R. Boia et al., “Logo Localization and Recognition in Natural Images using Homographic Class Graphs,” Mach. Vision Appl.,27(2):287–301, 2016.
3. K. Li et al., “Logo Detection with Extendibility and Discrimination,” Multimed Tools Appl., 72(2):1285–1310, 2014.
4. A. D. Bagdanov et al., “Trademark Matching and Retrieval in Sports Video Databases,” Proc. Workshop Multimedia Info.Retri., pp.79–86, 2007.
5. D.Delannay, F.Delaigle, H.Demarty and M.Barlaud, ”Compensation of Geometrical deformations for Watermark Extraction in Digital Cinema Applications”, in proc. of SPIE Electronic Imaging 2001, Security and Watermarking of Multimedia Content III, vol.4314, 149-157,2001.
6. S.Baudry, Bertrand Chupeau and F.Lef´bvreabc, ”Adaptive Video Fingerprints for Accurate Temporal Registration”, in proc. of IEEE Int.Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2010), pp. 1786–1789,2010.
7. D. Delannay, C. de Roover and B. Macq, ”Temporal alignment of video sequences for watermarking”, IS&T/SPIE’s 15th Annual Symp. on Elect. Imaging, California, USA, Proc. Vol. 5020, pp. 481-492, January 2003.
8. S.Baudry, B.Chupeau and F. Lef´bvreabc, ”A framework for video forensics based on local and temporal fingerprints”, in proc. of IEEE International Conference on Image Processing (ICIP 2009), pp. 2889–2892, 2009.
9. A.Saracog˘lu, E.Esen, T.K.Ates¸, B.O.Acar, Zubari, E.C.Ozan, E. o¨zalp, A.A.Alatan, and T. C¸ iloglu, ”Content Based Copy Detection with Coarse Audio-Visual Fingerprints”, in proc. of 7th Int. Workshop on Content-Based Multimedia Indexing (CBMI),213-218, 2009.
10. A. Divakaran, R. Regunathan, and K. A. Peker, ”Video summarization using descriptors of motion activity: A motion activity based approach to key-frame extraction from video shots,” Journal of Elect. Imaging, vol. 10, pp. 909-916,2001.
11. Sofia Tsekeridou and Ioannis Pitas, ”Content-Based Video Parsing and Indexing Based on AudioVisual Interaction”, in proc. of IEEE Trans. on Circuits & Sys. for Video Tech., Vol. 11, No.4,2001.