International Journal of Control and Automation
Vol. 13, No. 2, (2020), pp. 63-73

TOWARDS REAL TIME LOGO DETECTION AND CLASSIFICATION USING DEEP LEARNING

Kiran Kumar JP
Sri Siddhartha Institute of Technology (SSIT), SSAHE, Tumkur, India

MC Supriya
Sri Siddhartha Institute of Technology (SSIT), SSAHE, Tumkur, India

Abstract

In the recent years, there are significant changes in the methods used to solve the problems associated with logo detection, localization and classification. Deep learning has helped to achieve state of art results for logo detection, localization and classification. Key challenges with respect to logo detection network are to identify the location of the logo. Essentially the algorithms related to region proposal methods play a key role in categorizing the efficiency of logo detection networks. Some of the methods like SPPnet [1] and Fast RCNN [2] are efficient in bringing down the time required for detection, but it also indicates the overhead in terms of computation required for region proposal. This paper discusses about region proposal network (RPN) which enables and improves the efficiency of logo region proposals at real time. Our proposal includes method to extract the objects from content, feature selection and classification of the extracted object using convolutional neural network (CNN). Hence the proposed method consists of recognition pipeline which takes care of region proposal, logo classification based on specifically trained convolutional neural network.

Keywords*: Region proposal, Logo detection, Logo recognition, Convolutional neural network.*

INTRODUCTION

In recent years, we have seen the change in transmission of pay channel services. There is a paradigm shift from traditional mode of transmitting the services via cable, terrestrial, satellite to internet mode of transmitting the pay channel services. Considering the availability of device capabilities (like CPU processing capability, increased memory, enhanced display, etc.) these days, over the top (OTT) services are the most preferred ones. This is beneficial to the user because users can continue watching their favorite channels even when they are away from television. On the other side, care must be taken to ensure on protecting the content from piracy. Service provider, content owners will have a tough challenge in ensuring this.

Illegal redistribution of content is very much seen during high value event (like, world cup soccer, cricket, Olympics, etc.). The viewers watching across the globe will be more during these events. Typically, the content transmission between content delivery network (CDN) and end device (TV, Mobile, any portable device) will be encrypted based on using one of Digital Rights Management (DRM) technologies. Even if pirates eavesdrop the network and make a copy of this content, it's not possible to decrypt. Hence pirates might use the legitimate subscription for capturing the content.

Pirates make use of high-end camera for recording the content which gets decrypted at end device. The recorded content which is decrypted shall be subjected to minimal modifications, transcoding and shall be uploaded to streaming server. Popular streaming protocols (Real Time Messaging Protocol and Real Time Streaming Protocol) can be used for streaming the content. Easy way to publish / advertise these streaming URL's to end users would be through social networking websites (Twitter, Facebook, etc). Metadata may or may not be included during the retransmission of pirated content. However even if metadata is pushed along with pirated content, it's quite challenging to distinguish between original and pirated content.

ISSN: 2005-4297 IJCA

Popular protocols used to push pirated content can be categorized into direct web streaming and peer-to-peer methods. Direct web streaming is the most preferred method.

Figure1 illustrates how possible legitimate subscription can be used for capturing the content followed by retransmission.

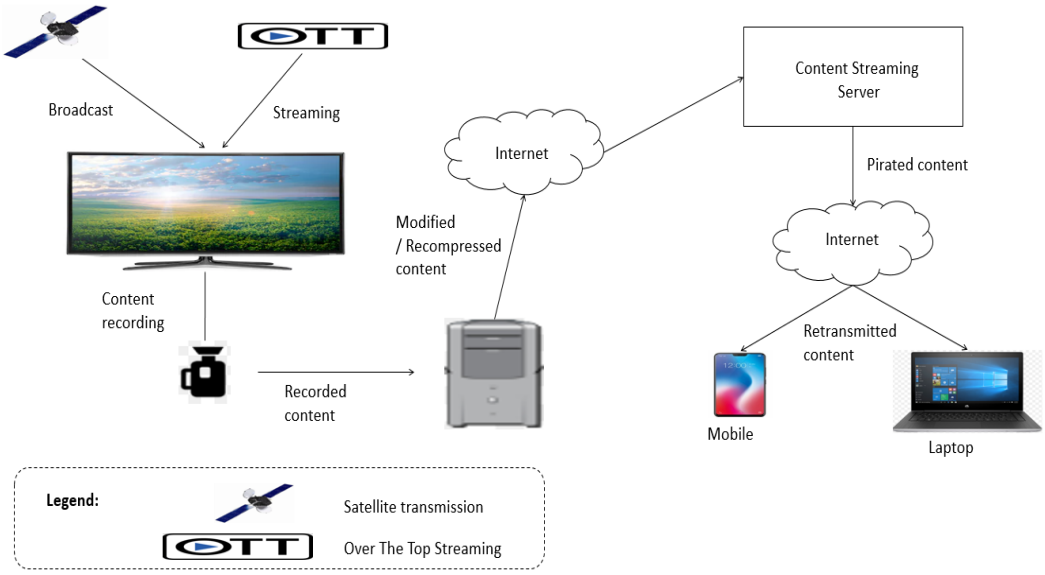


Figure1: Capturing and illegal retransmission of pirated content

To distinguish on actual or pirated content, video frames need to be analyzed. During high value events (like, sports), actual content can be transmitted through multiple physical channels (like, ESPN and star sports), this increases the difficulty and creates new challenges of analyzing video frames.

Basically, video frame analysis comprises of identification of visual objects, comparing these objects with original source content. The unique and comparable object extracted out of the video frames is broadcaster logo. This is the key identifier to detect and identify the pirated content. The key challenge here will be to locate and identify the broadcaster logo as the broadcaster logo would have been subjected to changes like occlusion, visibility, low quality, hidden during the retransmission of pirated content. Since the broadcaster logo has been changed up to some extent, this increases the difficulty level.

Artificial Intelligence (AI) solutions are growing and have created itself a unique space covering different problem statements. Artificial Intelligence plays a pivotal role in media industry as well. AI will be the potential candidate to automatically track illegal re-distribution of pirated content.

There shall be concerns raised by service providers, content owners to stop this piracy menace. Machine Learning (ML) provides the necessary framework to monitor the video frames continuously followed by detecting the broadcaster logo to distinguish between original and pirated content.

There is a need for logo detection and recognition in various business applications, such as automobile (vehicle logo detection), advertisements (product brand recognition), legacy broadcasting digital television and newest streaming services over the top (OTT) (Broadcaster logo classification) etc. Some of the common challenges that need to be addressed during logo detection and recognition include combination of graphics and text-based logos, modified / transformations like (scaling, translation and rotation) applied on the logos.

ISSN: 2005-4297 IJCA

Recently there has been lot of research explaining the usage of deep learning technique for detecting and classification of objects present in the video frames / Images. Deep learning technique is powered by convolutional neural network (CNN) method providing great accuracy for video recognition tasks. Below figure3 depicts the overview of simple deep neural network based on CNN to classify the object. Each layer has its own functionality to be performed. Certain neurons in each layer gets activated which in turn shall be the inputs for the neuron activation of next subsequent layers.

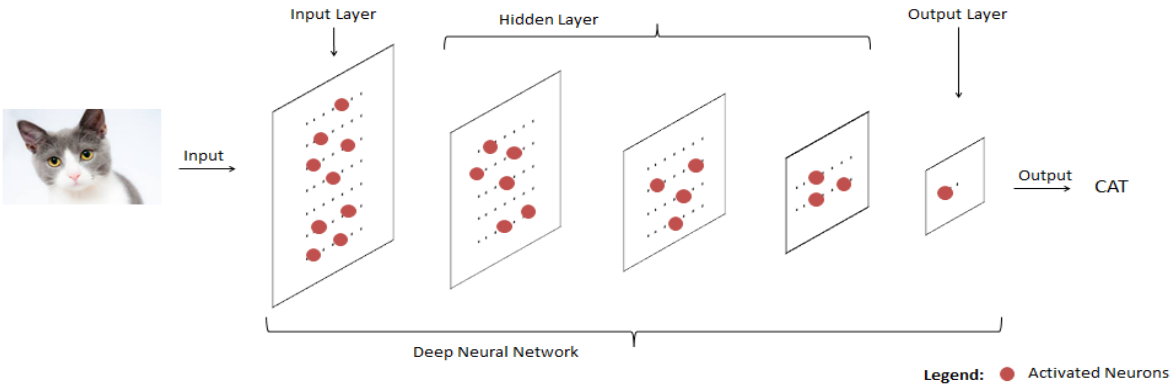


Figure3: Example depicting object classification based on deep neural network

One of the challenging tasks in object detection is to identify the intended availability of objects in the input frames. Once the object of our interest is available, it can be pushed for classification to know about the nature of the object. In this paper, we will discuss about the advantages of using region proposal network (RPN) over selective search object proposal methods. The subsequent sections shall elaborate more on this.

BACKGROUND

Real time identification and recognition of logo has become an essential task in many applications spread across different domain. Some of the key applications where logo detection and recognition is required is spread across shape and pose recognition for augmented reality [3], brand tracking from social media [4], vehicle logo identification for seamless traffic movement [5], content piracy detection, etc.

Legacy methods used for logo recognition are based on localization of key points, key point descriptor and matching. [6] Illustrates on categorizing the image into collection of local features, further extracting the score based on positive and negative queries. [7] Explains about the technique contrario threshold content-based extraction framework providing effective visual query methods. [8] Explains on data mining through spatial pyramid to extract local SIFT feature spatial configurations. [9] Proposes a method to identify trademarks appearing in videos based on SIFT descriptors. [10] Proposes a method to extract key points in the image based on Gaussian affine difference through elliptical features of SIFT. Further identification of logo and recognition is done by analyzing homographic graphs and creating class prototypes. Similar method is proposed in [11] which exploits SIFT keypoints of logo available in similar class and matching them through homographic graphs. [12] Proposes a statistical learning burstiness model for identifying the logo. A set of keypoints represents each logo image. Based on this learning, the weights will be lowered for incorrect logo detection. [13] Exploited an approach by analyzing local features identified in logo images and its relative spatial layout like edges and triangles, computing quantized representation of logo regions and reducing false positives. [14] Proposes a method based on feature bundling for recognizing the

ISSN: 2005-4297 IJCA

logo. Local features are clubbed with their spatial neighborhood features depicting more information about the content and, hence retrieving lesser false positives.

Recent researches in Machine Learning (ML) have shown that neural networks provide great accuracy for video recognition tasks. Neural network can be designed and trained for different visual recognition tasks like image classification, localization and detection. Neural network learns to distinguish objects in data without human intervention. [15] Proposed a method based on convolutional neural network instead of legacy key point-based techniques. Logo regions are detected based on segmentation algorithms followed by classification based on SVM using features computed by CNN. [16] Exploits a method using pre-trained CNN for identifying the regions and using SVM to classify these regions.

[17] Investigated several deep convolutional neural network (DCNN) architecture for logo recognition task. [18] Explains about using fast region-based convolutional neural network (FRCNN) for logo detection and recognition. CNN models pertained with ILSVRC Image Net data is used as input for this method. [19] Proposes simple and efficient identification algorithm for object detection. The algorithm has two parts, first applying CNN to localize and identify objects, second classification based on supervised pre-training for a specialized task.

[20] Proposes deep learning-based logo recognition using logo region detection algorithms followed by logo region classification using specifically trained convolutional neural network (CNN). [21] Introduces the method region proposal network (RPN). Region proposals being the bottleneck and hindering the overall performance of the recognition pipeline, RPN enables generation of quick and high-quality regions.

PROPOSED SYSTEM

Figure 4 describes proposed method for logo detection and recognition task. Considering that logo can appear in any corner of the image with or without scaling, orientation, regions which are more likely to contain objects are extracted through region proposal network. These proposed regions are wrapped as per the input dimensions required by fully connected layers through region of interest pooling.

Ensuring that the system should have great accuracy and performance, region proposal network (RPN) to generate region proposals is included in the pipeline (Figure 4). The CNN classifier shall be trained to classify the region proposals pushed out from region proposal network considering the regions proposed may or may not contain the logo.

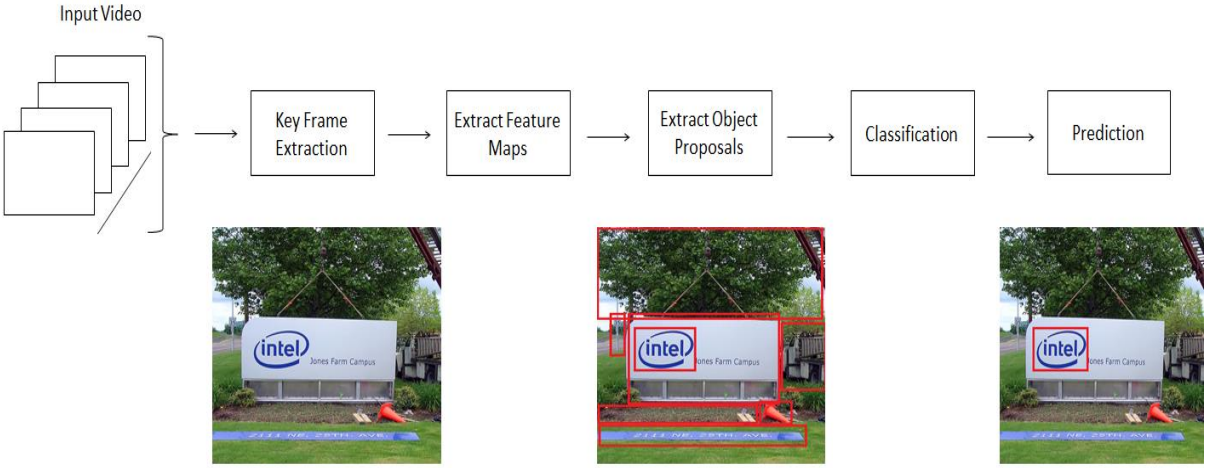


Figure4: Logo detection and recognition pipeline

The below figure 5 indicates overview of high-level network architecture used for learning and inference phase.

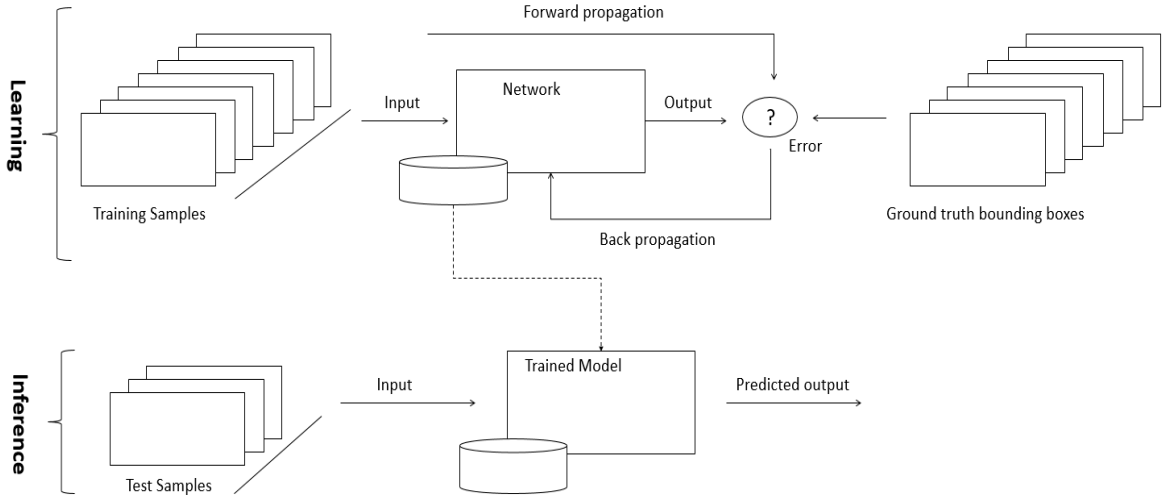


Figure5: Learning and Inference framework

The network block seen in figure5 is further elaborated in figure6. Figure 6 gives the detailed flow for learning phase.

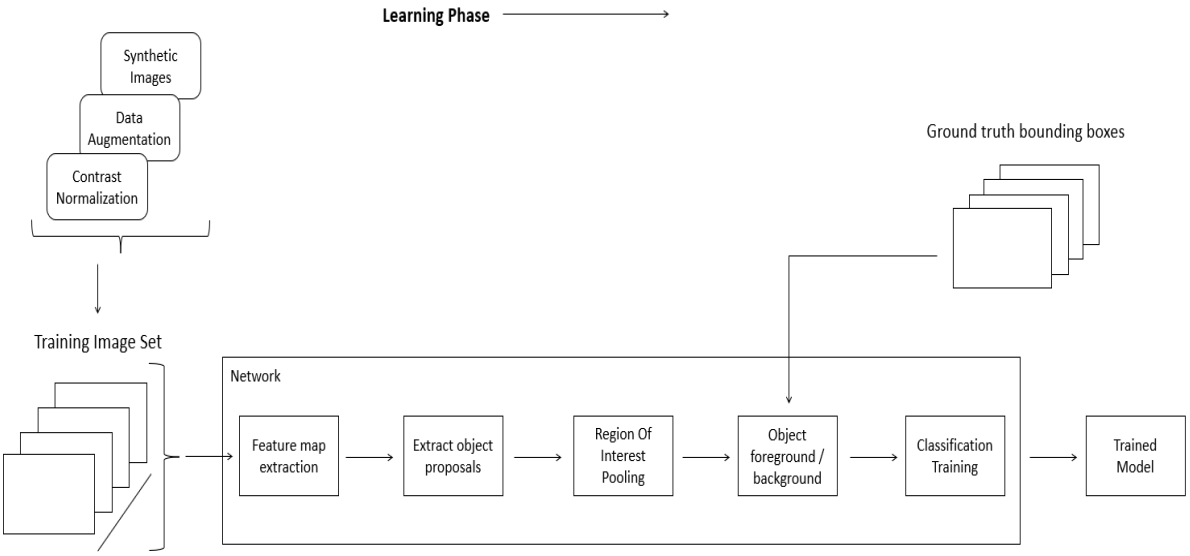


Figure 6: Learning phase - logo detection and classification.

Methods adopted to enhance logo detection:

Region proposals: Region proposal network (RPN) algorithm is used to extract regions that shall contain the logo. This algorithm is used in training and inference pipeline. The output of region proposal network gives two types of predictions: bounding box regression and binary classification. Proposed regions are considered as foreground if region proposed by RPN overlap with ground truth object with intersection of union (IoU) greater than 0.5. Proposed regions are considered as background if they don't overlap with ground truth object or intersection of union (IoU) is less than 0.1

ISSN: 2005-4297 IJCA

Ground truth logo annotation: The training image set undergoes marking the logo regions and hence annotate associating them with ground truth class. These marked logo regions may contain the combination of foreground and background, the reason being image logo shape and size.

Methods adopted to enhance logo classification:

Data augmentation: Training images are augmented to generate more data which can result in more sophisticated model. By doing this, it enhances the reliability and performance of CNN model and hence can be generalized.

Synthetic Images: Synthetic images are created by replicating the images of logo classes. This is to balance the logo for each class. By doing this, it enhances the CNN model classification and shall not be biased.

Contrast normalization: Method to prevent training images from having varying amount of contrast. This is done by subtracting the mean and dividing by standard deviation. By doing this, it enhances the CNN accuracy irrespective of image lighting conditions.

Key frame extraction:

In the sequence of video frames, extract key frames. Key frame (Intra-coded picture) is a complete image just like BMP or JPEG file. Key frames are coded without any reference to other frames except its own frames. These extracted key frames are pushed towards the downstream elements in the pipeline. Figure7 illustrates the sequence of I, B and P frames available in video stream sequence.

The input to the proposed pipeline is the continuous video stream of illegal transmission. Key frame is captured at regular interval of 20 seconds.

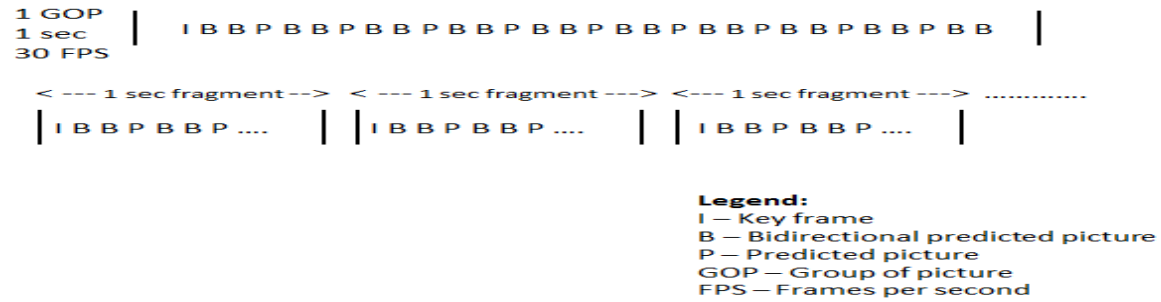


Figure7: I-frame, B-frame and P-frame available in video stream sequence

Region proposal extraction:

[21] Discuses about the state of art region proposal algorithm in object detection network. These algorithms implicitly improve the accuracy and performance of object detection to achieve state of art results.

This is a key module in the pipeline and its main responsibility is to extract region proposals in the given input image. Table1 explains the key difference between different techniques used for object detection and the region proposal methods used within.

Table1: Different methods of region proposals along with their limitations

Technique	Region proposal methods	Limitations
CNN	• Given image is divided into multiple regions. • Each region is further classified into different object class	• Accuracy can be improved only through availability of more multiple regions. • Breaking down to multiple region of interest is cumbersome. • Requires high computation
RCNN	• Given image is subjected to selective search algorithm to extract regions • Possibly extracts 2000+ regions for each image	• Expensive in terms of computational power. • Every region identified is pushed to CNN for classification and bounding box
Fast RCNN	• Given image is subjected to CNN once. • Feature maps are extracted. • Given image is also subjected to selective search (SS) algorithm to extract regions • Feature maps and regions extracted from selective search algorithm are combined to form wrapped image	• Requires high computation • Selective search algorithm is slow.
Faster RCNN	• Given image is subjected to CNN once to extract feature maps. • Features maps extracted are subjected to region extraction algorithm called Region Proposal Network (RPN) • Features maps extracted from CNN and regions proposals extracted from RPN are combined to form wrapped images	• Region proposal network is quite faster than selective search algorithm • The performance of RPN depends on the performance of CNN.

DATASET

Flickr logos 27 dataset

FlickrLogos-27 [22] is publicly available dataset consisting of 27 different logo brands with more than four thousand images. This is categorized into three sets:

1. Training set: Composed of 810 images associated to 27 logos classes/brands
2. Distractor set: Composed of 4207 logo images / classes
3. Test set: Composed of 270 images.

An example image depicting the classes of flickrlogos-27 is shown in Figure8

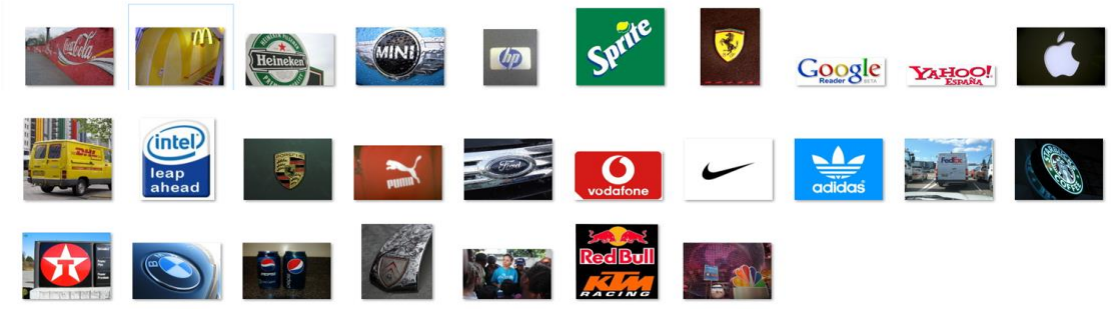


Figure 8: FlickrLogos-27 example showing images for each of 27 classes

Flickr logos 47 dataset

FlickrLogos-47 [23] is publicly available dataset consisting of 47 different logo brands with more than 8000 images. This has been designed to help evaluation of logo detection / recognition systems for real time images. The complete dataset is categorized into three sets:

1. Training set – Composed of totally 2769 images.
2. Distractor set – Composed of 3000 images.

ISSN: 2005-4297 IJCA

3. Test set – Composed of 1675 logo images.

An example image depicting the classes of flickrlogos-47 is shown in Figure9

Figure 9: FlickrLogos-47 example showing images for each of 47 classes

Experimental Result

System proposed in section3 is used for exploring and experimenting using the datasets mentioned in section 4 (flickrlogos-27 and flickrlogos-47). To enhance and improve the accuracy of the system, default images have been subjected to contrast normalization, data augmentation and even synthetic images are created. For object detection tasks, popular metric to precision and recall has been used to evaluate our model.

Prediction of the object of interest increases as the overlap region (prediction) is closer to ground truth. Figure 10 illustrates common understanding about intersection of union, ground truth and prediction regions.

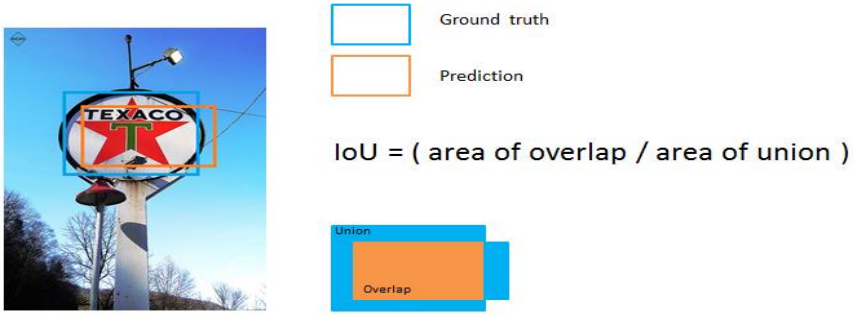


Figure 10: Ground truth and Prediction

Precision and Recall has been calculated as
Precision = TP / (TP + FP)
Recall = TP / (TP + FN) ---------- (1)
Where TP – True Positive, FP – False Positive, FN – False Negative

Precision and Recall has been calculated as updated in (1) for each class of flickrlogos-27 and flickrlogos-47. The table 2 and 3 indicates experimental results.
Proposal made by the system in section 3 is compared with ground truth to calculate intersection of union (IoU). Only the candidates whose IoU is greater than 0.5 is considered in below experiments. The confidence score is used to arrange the images.

Table2: Flickrlogos-27 precision and recall

Brand Logo	Precision	Recall
Adidas	0.87	0.62
Apple	0.96	0.82
BMW	0.74	0.84
Citroen	0.96	0.84
Cocacola	0.87	0.62
DHL	0.65	0.62
Fedex	0.96	0.86
Ferrari	0.91	0.86
Ford	0.91	0.82
Google	0.84	0.82
Heineken	0.96	0.84
HP	0.96	0.64
Intel	0.96	0.82

Brand Logo	Precision	Recall
McDonald	0.80	0.66
Mini	0.96	0.88
NBC	0.95	0.92
Nike	0.96	0.82
Pepsi	0.70	0.62
Porsche	0.84	0.88
Puma	0.84	0.88
Redbull	0.95	0.667
Sprite	0.95	0.94
Starbucks	0.91	0.84
Texaco	0.74	0.82
Unicef	0.91	0.82
Vodafone	0.95	0.92
Yahoo	0.58	0.66

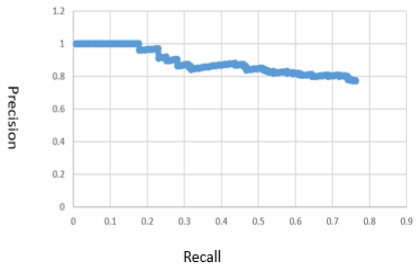
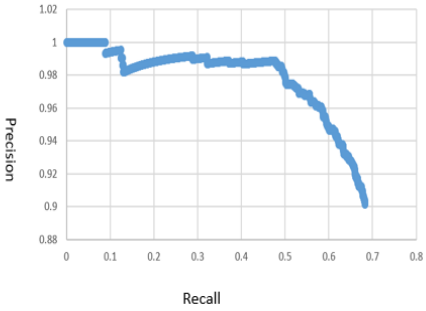


Table3: Flickrlogos-47 precision and recall

Brand Logo	Precision	Recall
Adidas	0.78	0.4
Aldi	0.98	0.68
Apple	0.98	0.62
Becks	0.99	0.75
BMW	0.99	0.68
Carlsberg	0.99	0.78
Carona	0.92	0.72
Chimay	0.99	0.866
Cocacola	0.97	0.58
DHL	0.91	0.48
Erdinger	0.98	0.73
Esso	0.95	0.56
Fedex	0.99	0.66
Ferrari	0.98	0.66
Ford	0.99	0.62
Fosters	0.98	0.74

Brand Logo	Precision	Recall
Google	0.96	0.63
Guinness	0.98	0.8125
Heineken	0.81	0.45
HP	0.96	0.45
Milka	0.99	0.8
Nvidia	0.78	0.45
Paulaner	0.97	0.74
Pepsi	0.96	0.7
Ritter Sport	0.99	0.8235
Shell	0.97	0.52
Singha	0.88	0.68
Starbucks	0.99	0.64
Stellaartois	0.95	0.84
Texaco	0.97	0.72
Tsingtao	0.98	0.74
UPS	0.96	0.56



Comparative Analysis:

This section focusses on comparing the current work with similar available work. Table4 indicates the comparison report.

Table4: Comparative study

Brand Logo	[17] AlexNet Precision	[17] VGG16 Precision	Ours Precision
Adidas	0.47	0.61	0.78
Aldi	0.69	0.67	0.98
Apple	0.68	0.84	0.98
Becks	0.71	0.72	0.99
BMW	0.81	0.70	0.99
Carlsberg	0.59	0.49	0.99
Carona	0.90	0.71	0.92
Chimay	0.70	0.33	0.99
Cocacola	0.58	0.92	0.97
DHL	0.56	0.53	0.91
Erdinger	0.83	0.80	0.98
Esso	0.89	0.88	0.95
Fedex	0.70	0.61	0.99
Ferrari	0.88	0.90	0.98
Ford	0.85	0.84	0.99
Fosters	0.86	0.79	0.98

Brand Logo	[17] AlexNet Precision	[17] VGG16 Precision	Ours Precision
Google	0.90	0.85	0.96
Guinness	0.81	0.89	0.98
Heineken	0.66	0.57	0.81
HP	NA	NA	0.96
Milka	0.46	0.34	0.99
Nvidia	0.52	0.50	0.78
Paulaner	0.98	0.98	0.97
Pepsi	0.42	0.34	0.96
Ritter Sport	0.63	0.63	0.99
Shell	0.50	0.57	0.97
Singha	0.83	0.94	0.88
Starbucks	0.99	0.95	0.99
Stellaartois	0.87	0.82	0.95
Texaco	0.81	0.87	0.97
Tsingtao	0.86	0.84	0.98
UPS	0.70	0.81	0.96

Discussions:

The proposed system has shown great result for predicting the objects of interest in-line with the ground truth proposals. Thus, out of the total object proposals made by this system across different classes is accurate and hence increasing the precision. However, in terms of recall, the system can be further enhanced to achieve state of art results for object proposals for the corner cases of occlusion.

CONCLUSION

This paper discusses about system architecture, methods used to enhance logo detection and logo classification using available generic data set (flickrlogos-27 and flickrlogos-47). This paper proposes the recognition pipeline with key feature as region proposal network (RPN) to extract the objects. Region proposal network is very much essential to reduce the computational time taken to generate object proposals of interest. We understand that position, size, orientation of the logos can vary differently and their appearance need not be constant. Further, logos might be explicitly distorted. Our proposed recognition pipeline includes classification that is specifically trained to recognize the logos for real time scenarios.

As future enhancements, dataset consisting of logos from different broadcasting content and over the top (OTT) services shall be used for learning and inference phase using the proposed architecture. Proposed system can be used to explore on newer object detection techniques like YOLO (you only look once) and SSD (single shot detector).

REFERENCES

[1] K. He, X. Zhang, S. Ren, and J. Sun (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition in European Conference on Computer Vision (ECCV).
[2] R. Girshick. (2015). Fast R-CNN. IEEE International Conference on Computer Vision (ICCV).
[3] N. Hagbi, O. Bergig, J. El-Sana, M. Billinghurst. (2011). Shape recognition and pose estimation for mobile augmented reality. IEEE Trans. Vis. Computer Graph. 17 (10) 1369–1379.

[4] Y. Gao, F. Wang, H. Luan, T.-S. Chua. (2014). Brand data gathering from live social media streams. International Conference on Multimedia Retrieval, ACM, p. 169.
[5] A.P. Psyllos, C.-N.E. Anagnostopoulos, E. Kayafas. (2010). Vehicle logo recognition using a sift-based enhanced matching scheme. IEEE Trans. Intell. Transp. Syst. 11 (2) 322–328.
[6] J. Meng, J. Yuan, Y. Jiang, N. Narasimhan, V. Vasudevan, Y. Wu. (2010). Interactive visual object search through mutual information maximization. ACM International Conference on Multimedia, ACM, pp. 1147–1150.
[7] A. Joly, O. Buisson. (2009). Logo retrieval with a contrario visual query expansion. ACM International Conference on Multimedia, ACM, pp. 581–584.
[8] J. Kleban, X. Xie, W.-Y. Ma, Spatial pyramid mining for logo detection in natural scenes, in: IEEE International Conference on Multimedia and Expo, 2008, IEEE, 2008, pp. 1077–1080.
[9] A.D. Bagdanov, L. Ballan, M. Bertini, A. Del Bimbo. (2007). Trademark matching and retrieval in sports video databases. International Work- shop on Multimedia Information Retrieval, ACM, pp. 79–86.
[10] R. Boia, C. Florea, L. Florea. (2015). Elliptical shift agglomeration in class prototype for logo detection. BMVC. 115–1
[11] R. Boia, C. Florea, L. Florea, R. Dogaru. (2016) Logo localization and recognition in natural images using homographic class graphs, Mach. Vis. Appl. 27 (2) 287–301.
[12] J. Revaud, M. Douze, C. Schmid. (2012). Correlation-based burstiness for logo retrieval. ACM International Conference on Multimedia, ACM, pp. 965–968.
[13] S. Romberg, L.G. Pueyo, R. Lienhart, R. Van Zwol. (2011) Scalable logo recognition in real-world images. ACM International Conference on Multimedia Retrieval, ACM, p. 25.
[14] S. Romberg, R. Lienhart. (2013). Bundle min-hashing for logo recognition. ACM Conference on International Conference on Multimedia Retrieval, ACM, pp. 113–120.
[15] S. Bianco, M. Buzzelli, D. Mazzini, R. Schettini. (2015). Logo recognition using CNN features, in: Image Analysis and ProcessingI–CIAP, Springer, pp. 438–448.
[16] C. Eggert, A. Winschel, R. Lienhart. (2015). On the benefit of synthetic data for company logo detection. Annual ACM Conference on Multimedia Conference, ACM, pp. 1283–1286.
[17] F.N. Iandola, A. Shen, P. Gao, K. Keutzer. (2015). Deep logo: hitting logo recognition with the deep neural network hammer, arXiv preprint arXiv: 1510.02131
[18] G. Oliveira, X. Frazão, A. Pimentel, B. Ribeiro. (2016). Automatic graphic logo detection via fast region-based convolutional networks. International Joint Conference on Neural Networks (IJCNN), IEEE, pp. 985–991
[19] R. Girshick, J. Donahue, T. Darrell, J. Malik. (2016). Region-based convolutional networks for accurate object detection and segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 38 (1) 142–158
[20] Simone Bianco, Marco Buzzelli, Davide Mazzini, Raimondo Schettini. (2017). Deep learning for logo recognition, Elsevier, Neurocomputing volume 245 Issue C, pages 23-30
[21] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. (2016). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, https://arxiv.org/abs/1506.01497
[22] Y. Kalantidis, LG. Pueyo, M. Trevisiol, R. van Zwol, Y. Avrithis. (2011). Scalable Triangulation-based Logo Recognition. In Proceedings of ACM International Conference on Multimedia Retrieval (ICMR 2011), Trento, Italy.
[23] Stefan Romberg, Lluis Garcia Pueyo, Rainer Lienhart, Roelof van Zwol. (2011). Scalable Logo Recognition in Real-World Images. ACM International Conference on Multimedia Retrieval (ICMR11), Trento. Also, Technical Report, University of Augsburg, Institute of Computer Science.

ISSN: 2005-4297 IJCA