Journal of Computational Science and Intelligent Technologies
eISSN: 2582-9041
DOI: 10.53409/MNAA/JCSIT/e202102033444
Type: Research Article

A Comprehensive Logo Dataset for Deep Learning-Based Classification for Content Piracy

Kiran Kumar Jakkur Patalappa[1] and Supriya Maganahalli Chandramouli[2]

[1,2]Sri Siddhartha Institute of Technology (SSIT), Sri Siddhartha Academy of Higher Education (Deemed to be University), Tumkur 572105, Karnataka, India.

[*]**Corresponding Author: Supriya Maganahalli Chandramouli. Email: supriya.mc9@gmail.com**

Received: 21 March 2020; Accepted: 10 July 2021

Abstract: Given the online infrastructure we have today, content piracy is growing and spreading quickly across various countries. These internet infrastructures' main goal is to offer a platform for delivering permitted and lawful material from the service provider to the end user. Over time, pirates have used the digital online infrastructure system to duplicate and retransmit the original content using the same infrastructure. Visual analytics of the broadcast logo is one way to determine if the content is pirated. The construction of a new scalable TV broadcast channel logo corpus spanning diverse geographies and genres, as well as the publically available datasets of TV broadcast channel logos (Indian channels), will be covered in this work. A total of 450 TV broadcast channel logos in various regional languages have been gathered for genres including (Sports, Movies, Kids and Cartoons, Entertainment etc.) Each logo is exposed to various data augmentation approaches to increase the logo corpus and boost the deep learning logo classification. This logo corpus with the cutting-edge object identification algorithm YOLO v2 is also covered in this study, along with the recognition of several logo classes. Results from experiments are documented for various inference logos with various pixel contexts.

Keywords: Data augmentation; Deep learning; Object detection; TV broadcast; Channel logo corpus; YOLO.

1 Introduction

In computer vision and pattern recognition, logo detection and classification have been the key research areas. Logo detection talks about identifying the location of the desired logo in the given input image. This is challenging as the desired logo of interest needs to be narrowed among different objects in the given input image. Logo classification talks about categorising the identified logo into a certain logo name. A wide range of applications spread across different domains where this logo detection and classification are applied [1]. Product recognition [2], vehicle logo detection and classification, social media brand image monitoring [3], copyright infringement identification, and recommendations for TV broadcast channels. The composition of logo objects will be a combination of text and graphic symbols. The logo contained in the real image can vary based on factors like size, shape, texture, transformations, rotation, contrast and occlusion. Hence the complexity of logo detection and classification in these cases will eventually increase.

Content piracy is causing a big impact on broadcasters and content providers, decreasing the value of

live programs (Like sports and entertainment). Considering the demand for this online content, there is an increase in personalised online content allowing users to choose their programs of interest. Because of huge demand, this will open the doors for pirates to act and work on content piracy and retransmit the content illegally. Figure 1 provides a high-level diagram of the flow of illegal content that pirates use. This includes setting up a hosting site where the address to stream transcoded media segments is placed, followed by a linking site that publishes and attracts users for access to content piracy. Pirates use powerful social media platforms to publish pirate links. One of the other methods to publish these pirate links could be through google ads. Using these pirate links, the re-encoded pirate content can be streamed at a cheaper price and thereby signalling a serious threat to the business related to content generation and distribution (Praveen, 2020).

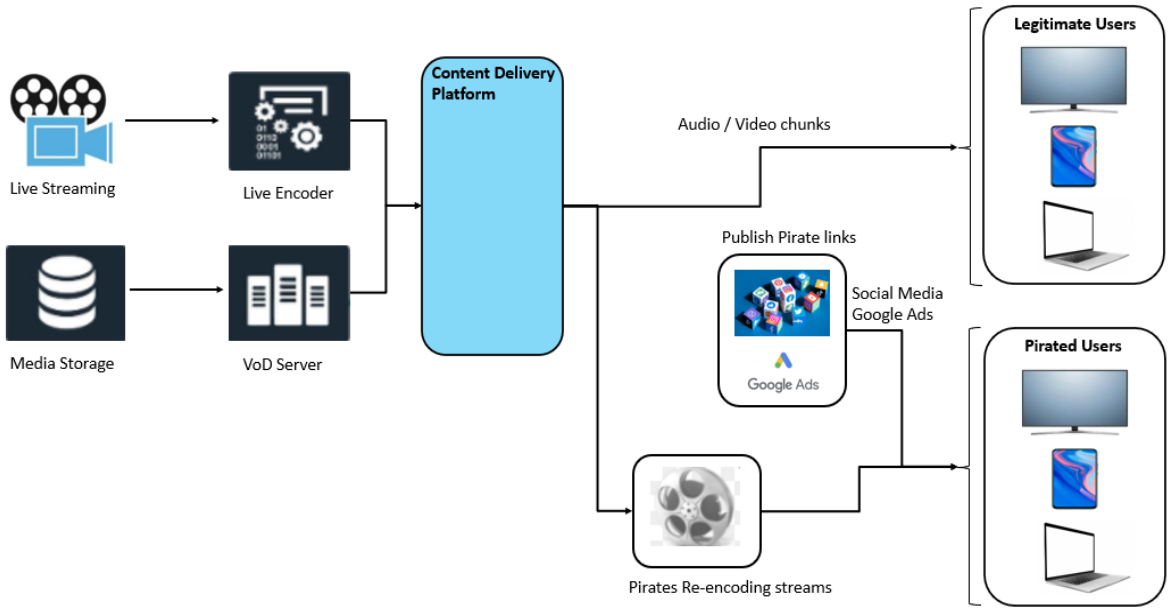

Figure 1: Flow of Pirated Content

The most important step towards curbing pirated transmission is to detect and identify that the transmitted content is pirated. This process can be categorised into the following steps, detection, collection, computation and updation. The first step is to detect the published links which point to pirated content. This is done through indexing web links and advertisements published on social media. The second step is to collect the streaming data, which needs to be analysed to check the presence of watermarking, forensic fingerprint or visual analytics of the content. The third step is to apply computations to the collected data. This could be through any of the different video analytics methods. The last step is to use the results made out of step 3 and take appropriate action. Visual analytics of the content becomes essential to compare the original content with illegal retransmitted content and identify content piracy in real-time.

The key identifier would be a broadcast logo. A broadcast logo is available in all video frames of the original content. Identification of broadcast logos can be handled independently without the need to integrate with the content delivery system and the content distribution system. Video frames need to be analysed for the placeholder of the broadcast logo and to detect the logo. These broadcast logos could have been distorted, modified, made visibly poor or hidden by the pirates in the retransmitted content. In this paper, we discuss the creation of the broadcast logo dataset comprising 450 logo classes spread across different regions in India. The creation of this large-scale dataset is expensive, challenging and time-consuming. Apart from the dataset, this research explores the state-of-the-art deep learning methods for generic object detection and uses this method for broadcasting logo detection and classification and evaluating extensive experiments.

2 Materials and Methods

2.1. Generic Object Detection

There are different generic object detection methods. The main objective of these methods is to locate the object, classify them in the given image, draw a rectangular bounding box marking the object's existence and associate the confidence score. The functioning of generic object detection methods can be broadly classified into two categories. Type 1 comprises of detection and classification pipeline, which includes extraction of region proposals followed by subjecting each region proposal to classification for different object categories. Type 2 comprises a unified approach for object detection and classification, categorising object detection as a regression or classification issue [4-16].

2.2. Region Proposal-Based Method

Figure 2 illustrates the R-CNN data flow diagram, which comprises Input Image processing, region extraction, Feature generation and classification. R-CNN uses a selective search algorithm to extract 2000 region proposals for a given image. Selective search algorithm groups similar regions based on size, shape, colour, texture etc.

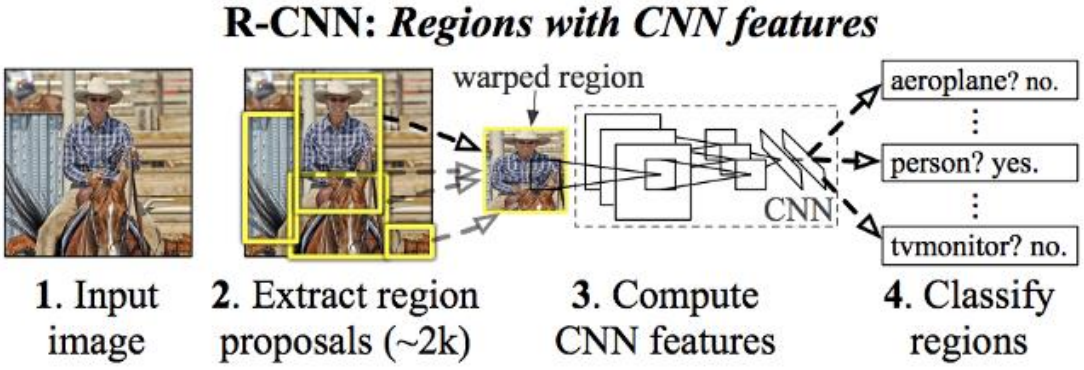

Figure 2: Data flow diagram of R-CNN [4].

2.3. Regression-Classification-Based Method

Region proposal-based methods include different steps in their pipeline, which are correlated for each step, from extracting proposals to gene generating. As a result, each step requires time to generate the results, creating a bottleneck for the real-time applications. Regression classification methods focus on generating bounding box coordinates and class probabilities based on the image pixels, thereby reducing the time required for classification. Figure 3 illustrates the data flow of SSD. The default ground truth bounding box available for each object in the input image shall be utilised in the SSD method. Each ground truth box is checked for varying aspect ratios. By doing this, shape coordinates and the confidence for all object categories available in the default box shall be predicted. During the training ground, truth boxes are compared with the default boxes, as shown in Figure 3. The matched ones shall be considered positive, while the rest are considered negative.

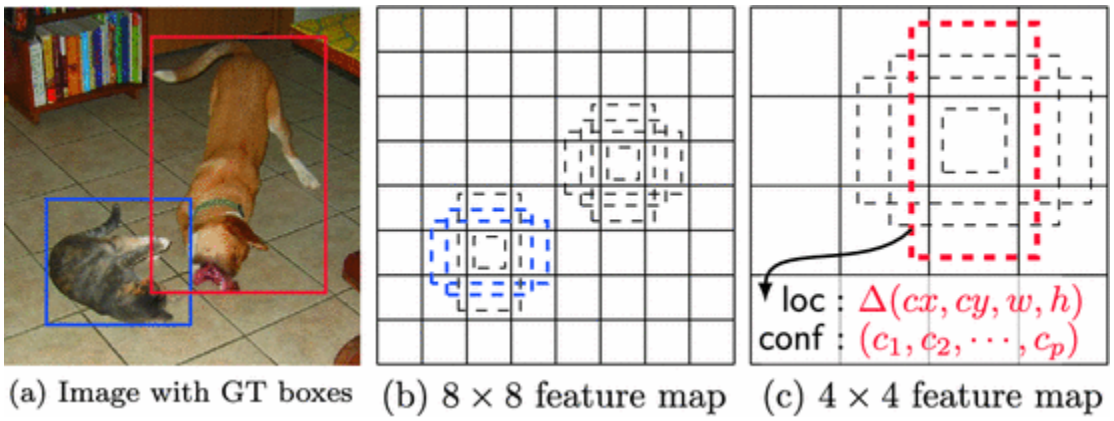

Figure 3: Single Shot Detector (SSD) method [10] includes an input image

2.4. Dataset Collection

This section will focus on dataset gathering from available sources, followed by expanding the available dataset using data augmentation techniques. Lastly, describing the different data augmentation techniques used. Publicly available TV broadcast channel logo datasets contain logos across different regions of the globe. These images are small in number and lack logo classes as well. It has been observed that the same logo class is quite repeated with variations in size and contrast.

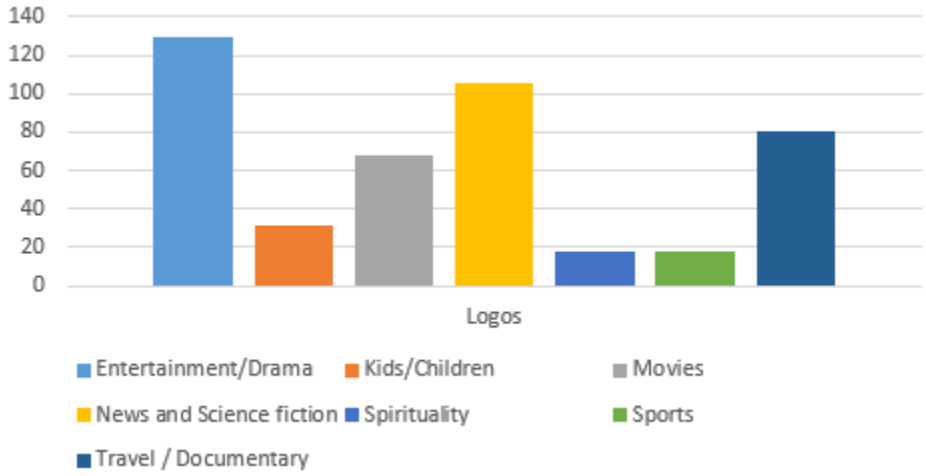

Figure 4: Images grouped as per their TV genre

In this paper, we have collected images of popular Indian TV broadcast channel logos. The logo of 450 TV broadcast channel logos has been collected using the google search engine, which is spread across different TV genres (Entertainment/Drama, Kids/Children, Movies, News and Science fiction, Spirituality, Sports, Travel/Documentary). Within each different TV genre category, care has been taken to pick the popular broadcast channel logo. Figure 4 indicates a pictorial representation of the logos captured and grouped under different TV genres, and Figure 5, updated below, captures a glimpse of TV broadcast channel logos collected.

Figure 5: Some popular Indian TV broadcast logo images retrieved from google search

Data augmentation is a technique to artificially increase the data size, thereby increasing the data diversity required for training the model without explicitly capturing new data. This technique focuses on expanding and enriching the depth of the data through transformations like cropping, titling, zooming etc. Figure 6 describes the original image being expanded/extended to more similar images with variations in the internal pixel values through various data augmentation techniques.

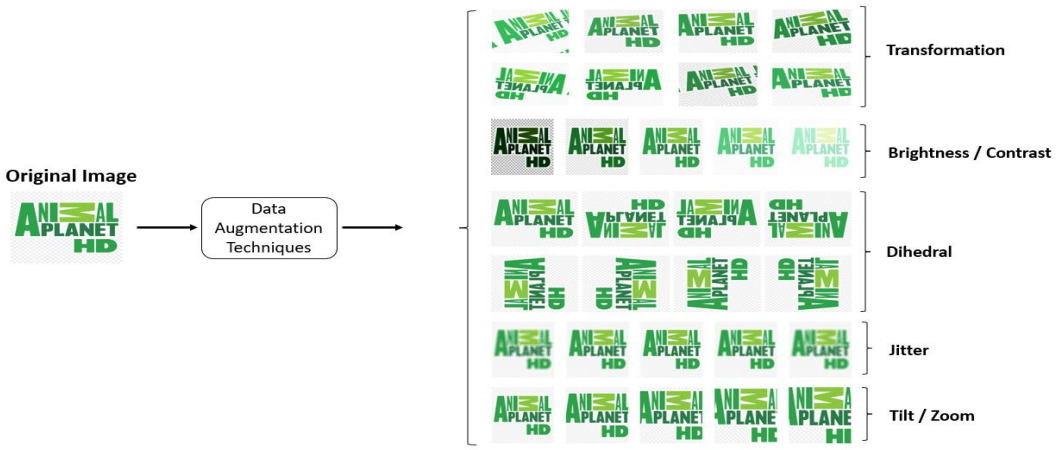

Figure 6: Some of the data augmentation techniques

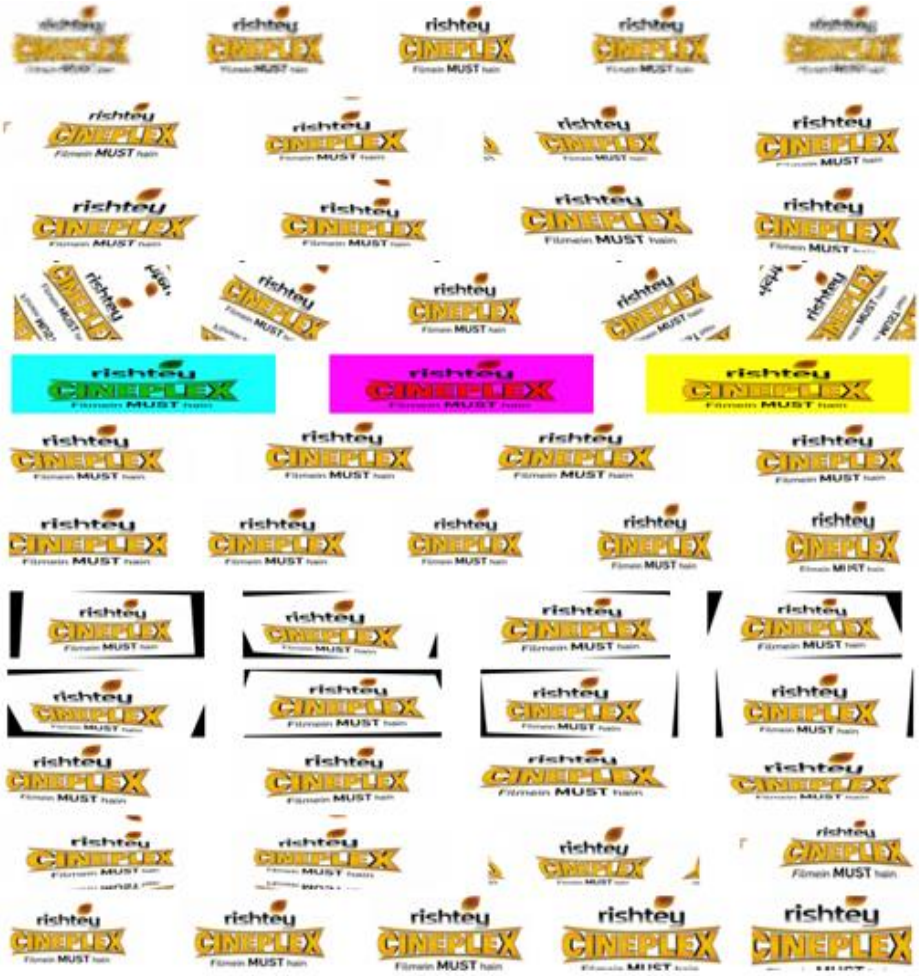

Figure 7: Transformed images after applying data augmentation techniques

Using the fastai image transformation library, the dataset collected, shown in Figure 4, has been expanded. Some of the key transformation applied includes (brightness, contrast, crop, dihedral, flip, jitter, perspective wrap, rotate, RGB randomise, skew, squish, symmetric wrap, tilt, zoom, and cut-outs). Figure 7 captures the transformed images after applying the fastai image transformation library.

3 Approach

Figure 8 illustrates the system diagram. In the training phase, the given input image is subjected to data augmentation methods to augment data of different kinds, eventually enhancing the classification model's performance and reliability. The accuracy of the classification model shall be enhanced even in cases where the logo is purposefully made with low visibility. The real data augmented images shall be the input to the block object detection based on the regression classification method. The real data augmented images are annotated, which shall be used by block object detection for training the model. In the testing phase, the system pipeline includes a key component, the regression classification method (YOLO), to extract different object classes for the given image. The objective of the block network of object detection and classification in the below diagram (figure 8) is to identify the objects running through image pixels, generate the coordinates around the identified object and predict the class probabilities.

Figure 8: System pipeline for extracting objects, classifying and bounding box

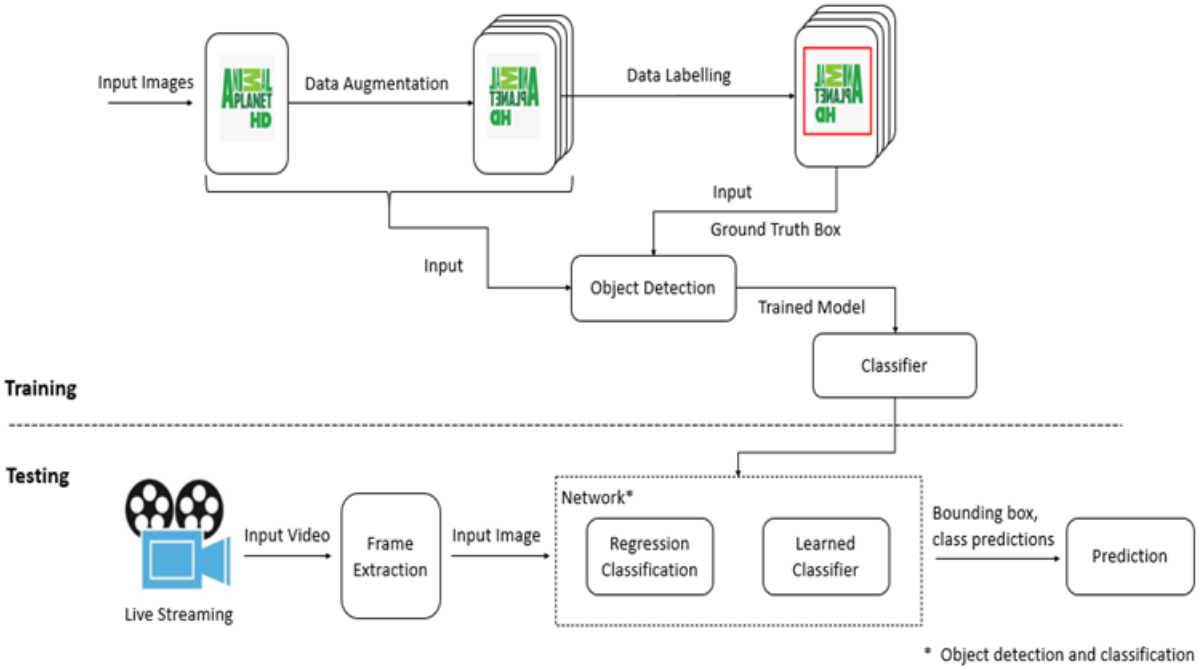

Figure 9 details the regression classification method in the context of a broadcast video image. Unlike region proposal-based object detection methods where each image is searched for the region of interest that could hold the object and then classify the identified object, regression classification method potentially targets to generate coordinates of bounding box and probability of class availability through image pixels in a single execution of the algorithm. The given input image is split into S x S grid of cells. The objects available in the image can be part of one S x S grid cell or spread across a collection of grid cells. The cell that hosts the object's centre and its corresponding grid is the grid responsible for predicting the object. Identification of objects and classifying them are done using a single CNN network. Every grid cell can predict B bounding boxes and Cp class probabilities. Bounding box prediction comprises 5 parameters (Bx, By, Bw, Bh, C). Bx, By corresponds to center of the box relative to a particular grid cell. Bw, Bh holds the width and height of the corresponding bounding box. Parameter C holds the confidence score about the availability of the object. The confidence score shall be zero if no object exists in the given cell. The

confidence score will be maximum when the intersection of the union (IoU) of the predicted box and the ground truth box are equal.

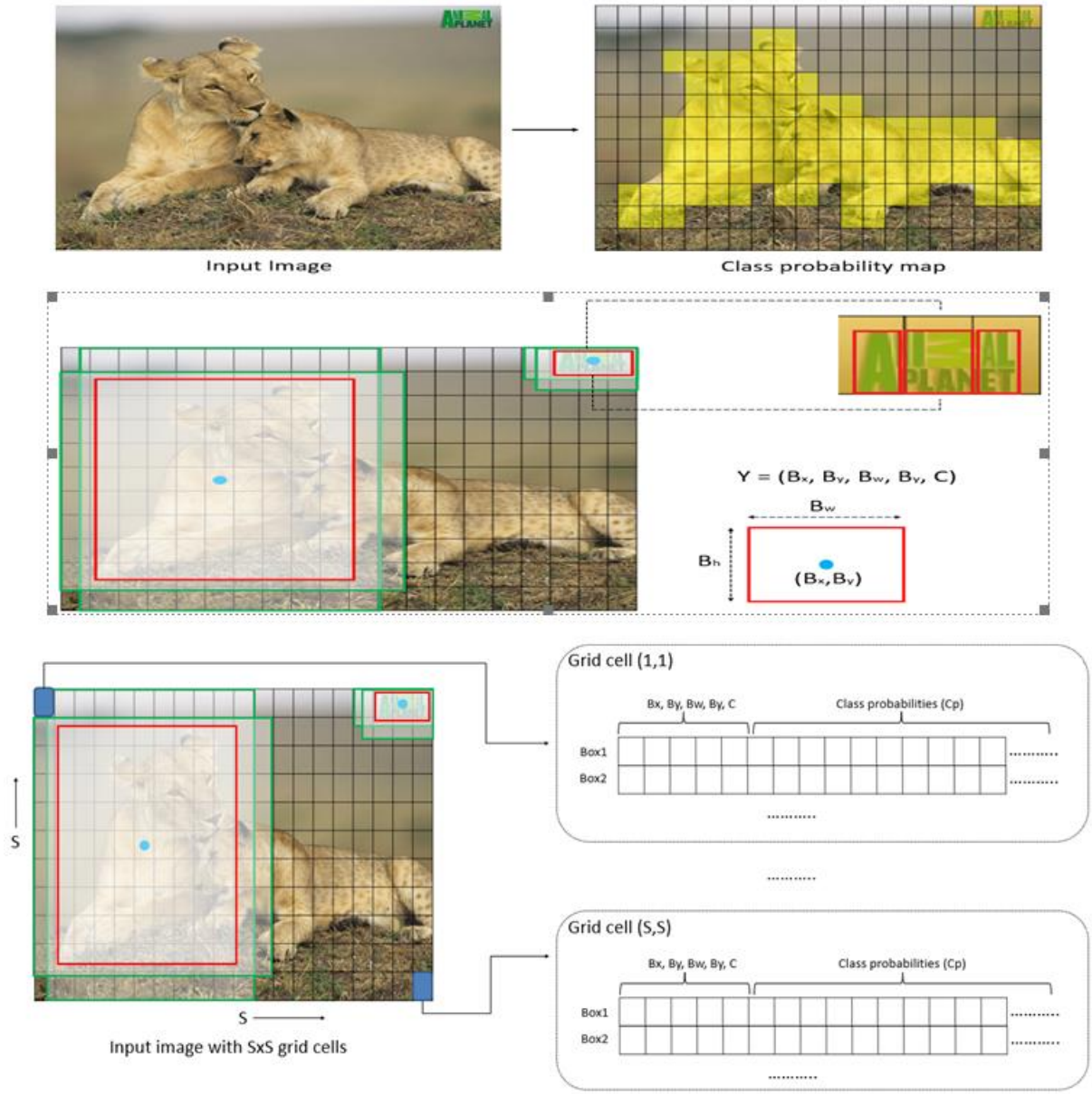

Figure 9: Object detection based on image pixel and classification

4 Results and Discussion

Each collected 450 different Broadcast TV channel logos (Figure 4) has been subjected to various data transformations, as shown in figure 6. Further, the data set is used on the system described in figure 7 for training the model. Each TV broadcast logo image is used as an input for different transformation methods to generate 125 real data augmentation images. Table 1 indicates the TV genre category, real data-augmented logos and total images after data augmentation techniques.

The complete dataset (Total images: 56250) is categorised into:

a. Training set – Consists of a total of 54000 images

b. Testing set – Consists of total of 2250 images

Table 1: Overall real data augmentation datasets

TV Genre	Logos	Augmentation Logos per TV broadcast logo	Total Images
Entertainment/Drama	130	125	16250
Kids/Children	31	125	3875
Movies	68	125	8500
News and Science Fiction	105	125	13125
Spirituality	18	125	2250
Sports	18	125	2250
Travel/Documentary	80	125	10000

Table 2 indicates the experimental results. Random images are picked from each TV genre category for verification using the trained model. For example, the TV genre category (Entertainment / Drama) consists of 130 different channel logo classes. Certain images are chosen from each different 130-channel logo class for verification. This is repeated for other TV genre categories. Precision, recall and accuracy have been calculated for each TV genre category.

The confidence score resulting from the intersection of union (IoU) of the predicted box and the ground truth above 0.5 is the targeted threshold for the below experiments.

Table 2: Channel logo experimental results

TV Genre	Logo Classes	Images used for Testing set	Precision	Recall	Accuracy
Entertainment/Drama	130	650	0.815	0.801	0.812
Kids/Children	31	155	0.823	0.794	0.803
Movies	68	340	0.807	0.781	0.801
News and Science Fiction	105	525	0.842	0.806	0.842
Spirituality	18	90	0.853	0.817	0.849
Sports	18	90	0.842	0.793	0.838
Travel/Documentary	80	400	0.832	0.806	0.831

Perhaps pirates will try to hide the logo through distortion or make them visibly poor. To artificially create this case, some data augmentation techniques such as random erasing/cut-out and jitter have been used. These artificially created images will mimic real-time behaviour about the pirated content, which has changed to the broadcast logo. The below picture in figure 10 captures some of the results taken from our trained model.

Figure 10: Results for random erasing/cut-out, jitter images

5 Conclusion

This paper focused on generating a broadcast video logo dataset, the mode to increase the data set through data augmentation techniques covering different image transformations. The idea behind the image transformation is to create artificial images that simulate a real environment where the pirates will makeover the logo while redistributing the pirated content. This paper discussed the different methods of generic object detection algorithms and system pipeline, which uses regression classification methods to identify the object and its location and classifies it. Experiments were carried out on different TV genre categories and the results obtained were promising. Systematically investigated the role of random erasing/cut-out jitter on the object recognition performance.

Acknowledgments: The authors thank Sri Siddhartha Institute of Technology (SSIT), Sri Siddhartha Academy of Higher Education (Deemed to be University), Tumkur, Karnataka, India for their continued support.

Funding Statement: The author(s) received no specific funding for this study.

Availability of Data and Materials: The data used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] H. Bin, Z. Nuoya, Z. Qiang, W. Xinggang and L. Wenyu, "DiffNet: A Learning to Compare Deep Network for Product Recognition," *IEEE Access*, vol. 8, no. 1, pp. 19336-19344, 2020. https://doi.org/10.1109/ACCESS.2020.2967090

[2] D. Erhan, C. Szegedy, A. Toshev and D. Anguelov, "Scalable Object Detection using Deep Neural Networks," *arXiv:1312.2249*, 2013. https://doi.org/10.48550/arXiv.1312.2249

[3] C. Fu, W. Liu, A. Ranga, A. Tyagi and A. Berg, "DSSD: Deconvolutional Single Shot Detector," arXiv:1701.06659, 2017. https://doi.org/10.48550/arXiv.1701.06659

[4] R. Girshick, "Fast R-CNN," *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 1440-1448, 2015. https://doi.org/10.1109/ICCV.2015.169

[5] K. He, X. Zhang, S. Ren and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916. https://doi.org/10.1109/TPAMI.2015.2389824

[6] H. Kaiming, G. Gkioxari, P. Dollar and R. Girshick, "Mask R-CNN," *In Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2980-2988, 2017. https://doi.org/10.48550/arXiv.1703.06870

[7] J. Dai, Y. Li, K. He and J. Sun, "R-FCN: Object Detection via Region-based Fully Convolutional Networks," *In Proceedings of the 30th International Conference on Neural Information Processing Systems*, Barcelona, Spain, pp. 379–387, 2016. https://doi.org/10.5555/3157096.3157139

[8] T. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan and S. J. Belongie, "Feature Pyramid Networks for Object Detection," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936-944, 2017. https://doi.org/10.1109/CVPR.2017.106

[9] L. Liu, D. Daria and M. Natalie, "Visual Listening In: Extracting Brand Image Portrayed on Social Media," Marketing Science, vol. 39, no. 4, pp. 1-18, 2018. https://doi.org/10.1287/mksc.2020.1226

[10] W. Liu, D. Anguelov, E. Dumitru, S. Christian, R. Scott *et al.*, SSD: Single Shot MultiBox Detector. *In proceedings of the 14th European Conference on Computer Vision – ECCV 2016*, Amsterdam, Netherlands, pp. 21-37, 2016. https://doi.org/10.1007/978-3-319-46448-0_2

[11] M. Najibi, M. Rastegari and L. S. Davis, "G-CNN: an iterative grid-based object detector," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Nevada, United States. pp. 2369-2377, 2016. http://dx.doi.org/10.1109/CVPR.2016.260

[12] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," *In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6517-6525, 2017, http://dx.doi.org/10.1109/CVPR.2017.690

[13] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *In Proceedings of the 28th International Conference on Neural Information Processing Systems*, Montreal, Canada, pp. 91-99, 2015. http://dx.doi.org/10.5555/2969239.2969250

[14] J. R. Uijlings, K. E. Van De Sande, T. Gevers and A. W. Smeulders, "Selective search for object recognition," *International Journal of Computer Vision*, vol. 104, no. 2, pp. 154–171, 2013. https://doi.org/10.1007/s11263-013-0620-5

[15] Z. Zhao, Z. Peng, S. Xu and X. Wu, "Object Detection with Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212-3232, 2019. http://dx.doi.org/10.1109/TNNLS.2018.2876865

[16] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, pp. 580-587, 2014. https://doi.org/10.1109/CVPR.2014.81