Providing brands visibility data in live sports videos using deep learning algorithms[*]

Julius Gudauskas[1,*,†]

Department of Applied Informatics, Kaunas University of Technology, Studentu 50, Kaunas, Lithuania

Abstract
In the dynamic landscape of marketing and advertising, assessing brand visibility in live sports events plays a pivotal role in understanding brand exposure and impact. Traditional methods of manual annotation and analysis are time- consuming and subjective, necessitating automated solutions for efficient and objective evaluation. In this study proposed a novel approach leveraging deep learning algorithms to evaluate brand visibility in live sports videos. This research employs state-of-the-art object detection models, such as YOLO (You Only Look Once) and Faster R-CNN, to detect and localize brand logos within video frames. By training these models on annotated open-source logo datasets, we can extract valuable insights about the brands. The experimental results demonstrate the effectiveness of the proposed methodology in detecting logos and providing a valuable data about the positions for brand owners.

Keywords 1
Brands visibility, logo detection, YOLO, Faster R-CNN

1. Introduction

In the world of sports live video streaming, a considerable number of brands are trying to get noticed by the audience using various popular visibility materials: posters, stickers, billboards, etc. The research evaluating the impact and effectiveness of advertisements in sports arenas emphasizes that people notice at least some of the advertisements they are exposed to and usually remembers a part of them that were the most noticeable [1]. Typically, clients engage in negotiations with advertising executives to determine the conditions that will govern brand placement in the arena – specifying factors like coverage, frequency of display on advertisement billboards, and overall visibility strategies. A study conducted by Eventmarketer
[2] found that 72% of the audience are captivated by the brand when they see it during the events like music festivals or sports competitions where they are provided with good emotions and excitement. These occasions, characterized by heightened emotions and excitement, offer a unique opportunity for brands to establish a connection with a vast and diverse audience, potentially converting them into new users. However, to reach a wider audience, the brand must be placed in a visible location. Studies has shown that locations, such as boundary line hoarding are considered as the perfect place to display brand logos without irritating the viewers and getting maximum visibility [3] [4]. As sponsorship agreements comes with a significant cost, brand owners are interested to know if their investment is paying off. But measuring the effectiveness of different brand placement can prove challenging, time consuming and requiring manual work, leaving brand owners uncertain about the true impact of their investment in sponsorship deals.

IVUS2024: Information Society and University Studies 2024, May 17, Kaunas, Lithuania
[1,*]Corresponding author
[†] These author contributed equally.
✉ julius.gudauskas@ktu.edu (J. Gudauskas)
🆔 0009-0009-6351-5080 (J. Gudauskas)
© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

A comprehensive understanding of the effectiveness of brand advertising through the integration of deep learning techniques in visual material analysis remains an evolving area, prompting the need for further research to refine methodologies and uncover insights that can inform strategic marketing decisions. To calculate the brand visibility metrics, the logos detection algorithm, that has the ability to detect all the logos, needs to be created. Current logo detection methods often focus on a limited set of logo categories, necessitating extensive training data that includes annotations for object bounding boxes [5]. But the main challenge remains the steep growth of the existing brand's amount and the brand image changes in the current ones. In this research, we addressed an open logo detection challenge and provided a unified brand's logo detection and recognition approach, using up to date machine learning algorithms.

2. Related works

Simple methods, that were developed as the early approaches for specific logo detections, relies on manually engineered visual features and traditional classification models [6] [7]. But such methods have a huge flaw: the region selection search algorithm based on sliding windows struggles to provide high accuracy in the high time complexity, manually created features lack robustness for logo diversity. Recent advancements in deep learning have revolutionized the field of visual material analysis, providing novel avenues for detecting and recognizing various objects, including logos. Many researchers have explored the application of deep neural networks in various image recognition tasks, demonstrating their capability to extract complex features and patterns. In the context of logo detection, the solution is usually determined by the size of the logos that has to be detected. Compared to a bigger size logos, smaller ones are more difficult to detect for several reasons: small logos with the low resolution contain little visual information, small logos covers a small area, it's surrounding box is more challenging to locate and there are usually less small logo samples [8]. In the small vehicles logo detection problem proceeding [9], researchers solved the issue using a YOLO [10] algorithm. In contrast to traditional methods relying on manual feature extraction, this system offers the benefits of self-learning features and direct image input. It can efficiently achieve both vehicle logo positioning and recognition functions. The researchers also introduce the Fast RCNN approach, that employs deep neural networks, utilizing convolutional layers to progressively extract abstract feature representations learned from previous convolutions [11] [12].

Natural visual scenes usually exhibit complexity and diversity - logos face various challenges, including object interference, shape distortion, different lighting, and limited perspective effects, which increases the difficulty of logo detection. In a recent development, researchers introduced a transfer learning approach, leveraging Densely Connected Convolutional Networks (DenseNet) for logo recognition [13]. They tested their method on the FlickerLogos-32 dataset and reached the accuracy higher than 92%. The visibility can also be impacted by the bad weather conditions. In the challenge, where the logos has to be detected in bad weather, authors presented an object proposal generation system AttentionMask [14]. The experimental findings indicate that the suggested approach demonstrates strong capabilities in identifying logos within intricate real-world settings. Nonetheless, data gathered from real-world scenarios may not match the quality of artificially augmented data, leading to a decline in the model's performance when detecting images in such real-world conditions.

3. Methodology

This work proposes a solution: a model that can identifying and predicting the bounding box around any logo within an image, irrespective of brand. Given the constant influx of new businesses, maintaining an up-to-date model with sufficient data for each brand proves challenging, potentially leading to difficulties in detecting logos of newly established brands during inference. To address this, the aim is to develop a model focused solely on detecting logos in images, regardless of brand affiliation. This approach eliminates the need to specifically train the model on individual brands, ensuring accurate detection of logos irrespective of their origin. The solution is focused on creating the model with highest accuracy, so the implementation is done using two different object detection algorithm pipelines: one – stage, that utilizes a single pass of the input image and enables the processing of the entire image in one go and two-stage models, that uses two passes of the image to make a prediction, where first pass is used to generate a set of proposals or potential object locations, and the second pass is used to refine these proposals and make final predictions [15]. In the category of one – stage models, YOLOv7 [16] visualized in Figure 1, were selected, because of its advantage in detecting smaller objects compared to a single – shot detection approach, and for two – stage - selected Faster R-CNN, that is visualized in Figure 2.
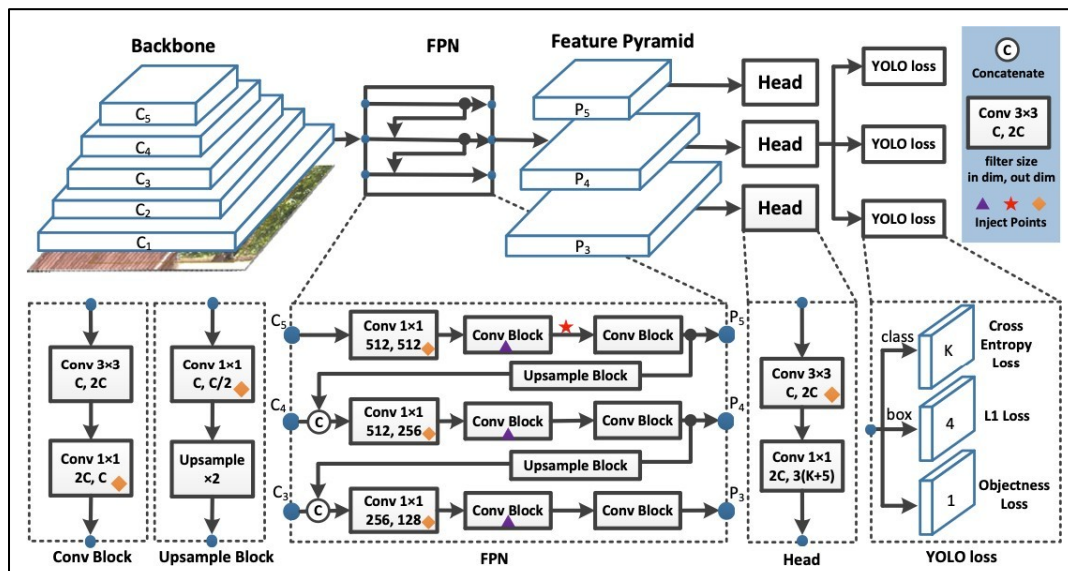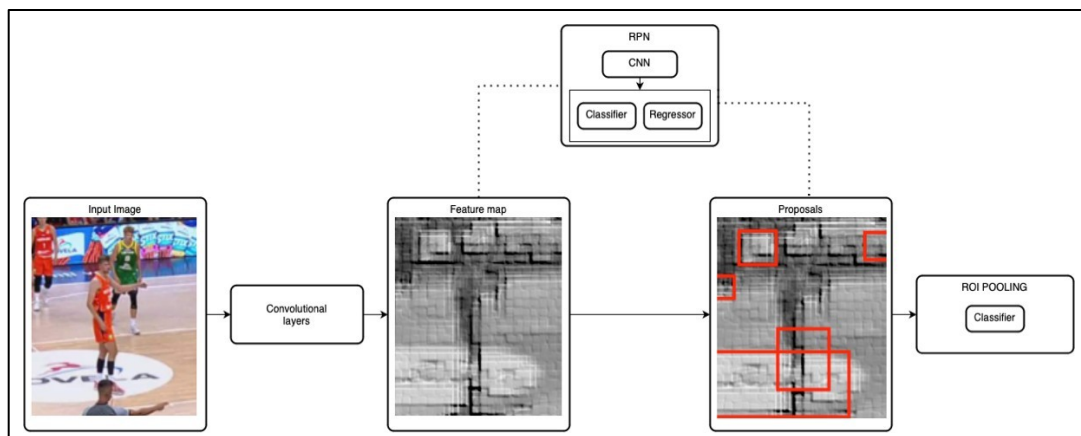

Figure 1: YOLOv7 workflow [16].

Figure 2: Faster R-CNN workflow.

3.1. Data

In conducting research on existing open-source logo datasets in Table 1, the attention was dedicated to comprehensively evaluating the diversity of logos within these repositories. Given the diverse nature of logos, ranging from graphic designs to text-based representations, the consideration was given to ensure that the selected datasets contain a wide spectrum of logo types.

Table 1
Open-source logo datasets.

Dataset name	Logos	Images	Year
BelaLogos [15]	37	10,000	2009
FlickrLogos-32 [16]	27	2240	2011
SportLogo [17]	31	2836	2020
QMUL-OpenLogo [18]	352	27,083	2018
FoodLogoDet-1500 [19]	1500	99,768	2021
LogoDet-3K [18]	3000	158,652	2020

FlickrLogos-32 [15]: This dataset includes 32 different logo classes from various domains. Since it contains about 2240 images with marked boundaries coordinates, it is well suited for building a model for brand detection and recognition (Table 2.)

Table 2
FlickrLogos-32 analysis.

Dataset part	Description	Images	Total images count
P1	Selected pictures with clean backgrounds. One brand logo in each image.	10 for each logo	320
P2	Different images showing at least one brand logo in different conditions. Pictures without logos.	30 for each logo 3000 without any logo	3960
P3	Different images showing at least one brand logo in different conditions. Pictures without logos.	30 for each logo 3000 without any logo	3960

QMUL-OpenLogo [18]. This dataset contains more than 27000 images with 352 different logos.It is a benchmark dataset for logo detection, formed by combining seven existing datasets and establishing an open protocol for evaluating detection performance. This dataset demonstrates a significant imbalance in distribution and notable variations in scale, crucial aspects for evaluating the effectiveness of detection algorithms.

LogoDet-3k [18]: This dataset contains more than 3000 unique classes, 158652 images with labeled logo symbols (dataset example in Figure 3.). This dataset divides logos into 9 different sub – categories: food, clothes, necessities, electronics, transportation, leisure time equipment, sports, and medicine and other (described in Table 3.). The main advantage of this data set is the large number of variations of the same brands – positions, lighting conditions, angles.
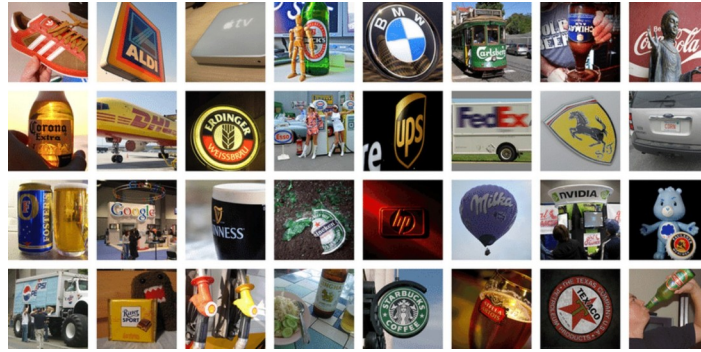
Figure 3: LogoDet-3k visualisation.

Table 3
LogoDet-3k analysis.

Category	Images count	Total brands count
Food brands	53,350	64,276
Clothes brands	31,266	37,601
Necessities brands	24,822	30,643
Electronics brands	9,675	12,139
Transportation brands	10,445	12,791
Leisure time equipment brands	5,685	6,573
Sports brands	3,953	5,041
Medicine brands	3,945	5,185

3.2. Data preprocessing

In object detection models, the training data encompasses crucial elements, such as the images itself, the coordinates of object bounding boxes, and their respective labels. Brand logo datasets commonly feature annotations tailored to individual brands – each brand is labeled by the name of it. However, this approach presents a notable challenge: models must be trained separately to detect each brand and will require additional fine – tuning to recognize newly introduced brands. Moreover, some brands will contain more training images than the others, so it also introduces a class disbalance problem that will potentially impact the model's performance.

In response to these challenges, introduced a data preprocessing step, that involves categorizing brands into two distinct groups: logos with textual elements and graphic – based logos. The result is achieved in this workflow visualized in Figure 4.:

- Each image is cropped by the bounding box coordinates.
- The image is processed using pytesseract – one of the most popular Python libraries for optical character recognition.
- If the optical characters were detected – assign the logo with the "TextLogo" label. If not – assign the logo with the "GraphicsLogo" label.
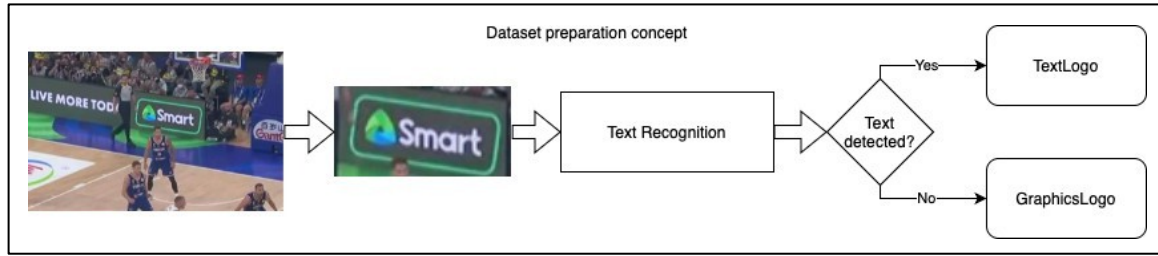
Figure 4: Data pre-processing pipeline

Rather than developing a new dataset, this approach leverages existing ones. Such approach not only saves time and resources, but also ensures that the model benefits from a diverse range of logo samples. Additionally, providing a minimum of two classes for training is essential for any object detection model to effectively learn and generalize. The final dataset for model training is stored in COCO (Common Objects in Context) format.

The newly created dataset (in Table 4.) contains a noticeable class imbalance This might have a huge impact in the model's ability to accurately assign the correct labels, leading to poor overall predictions. To address this issue, more graphic logos from other datasets were added, leaving the final dataset with a similar amount of each category. By ensuring a balanced distribution of samples across both categories, the model is equipped to learn effectively from a diverse range of examples, enhancing its capacity to make accurate predictions across all classes The final dataset was formed using the same amount graphics logos and text logos.

Table 4
Logo categories results after preprocessing.

Dataset	Text logo	Graphics logo
LogoDet-3K	168,416	25,845
FlickrLogos-32 and others	113	3,154
Total	281,416	29,008

4. Results

This research presents a comparative analysis between two object detection frameworks: YOLO (You Only Look Once) and Faster R-CNN. Specifically focusing on their efficiency to detect logo's bounding boxes, the aim is to provide an accuracy comparison between one-stage and two-stage detection methodologies. The comparison analysis is essential for understanding the trade-offs between accuracy and efficiency. While one-stage detectors are generally faster, they might sacrifice some accuracy compared to two-stage detectors. By quantitatively comparing the accuracy of bounding box detection between the two methods, determination can be done whether the sacrifice in accuracy is acceptable given the efficiency gains.

The experiments were done using specifically crafted dataset (3.2) with the model's parameters that provided the best accuracy results:

•   Faster R-CNN with ResNet-50 V1 FNP backbone, leveraging pre-trained weights, a learning rate (lr) set at 10-4, a momentum (m) of 0.9, and a weight decay (wd) of 10-6 and a batch size of 64.

•   YOLOv7 with a learning rate (lr) set at 10-4, L2 regularization at 10-4 and batch size of 64.

The evaluation of deep learning models was made using an mAP metric. Mean Average Precision (mAP) is a widely used metric for evaluating the performance of machine learning models, particularly in object detection and recognition tasks. It provides the assessment of a model's ability to accurately identify objects within an image dataset. The mAP metric

calculates the average precision for each class of objects across all images, then averages these values to produce a single score. This score reflects both the precision (the ratio of true positive predictions to all positive predictions) and recall (the ratio of true positive predictions to all actual positives) of the model. A higher mAP indicates better performance, with a score of 100% representing perfect detection accuracy. By utilizing mAP, we can quantify and compare the effectiveness of different models, aiding in the advancement of computer vision technologies.

Both model's data were augmented using random sized box crop with probability of 0.5, preserving the integrity of logos bounding boxes during random cropping and resizing. Horizontal Flip and Vertical Flip operations were applied independently, each with a 0.3 probability, introducing variations in viewpoint by flipping images horizontally and vertically. Data augmentation helped to increase YOLO model performance by 3% mAP and Faster

R-CNN by 2% mAP. Both model accuracies in the last 70 epochs presented in the chart Figure **5.** Even though, the Faster R-CNN achieved better accuracy results, the speed cost is significant compared both methods. During the performance analysis noticed, that YOLOv7 performs 13% faster. The Faster R-CNN model is more favorable when more time is available, but the YOLO algorithm is more favorable for the real-time task.
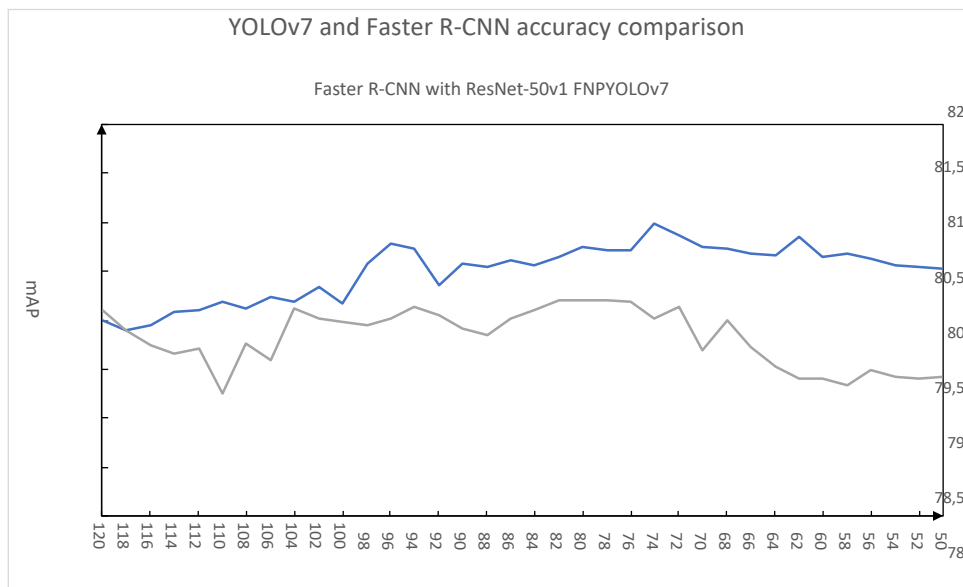

Figure 5: Faster R-CNN and YOLOv7 comparison.

During the comparison analysis between different Faster R-CNN FNP architectures noticed that ResNet-50 v1 demonstrates noticeable accuracy improvement over MobileNet-V3 and VGG16. The evaluation based on mAP metric shows that ResNet-50v1 provides up to 6% better accuracy results, compared to MobileNet and up to 5% compared to VGG16. Different FNP accuracies in the last 70 epochs presented in chart Figure 6.
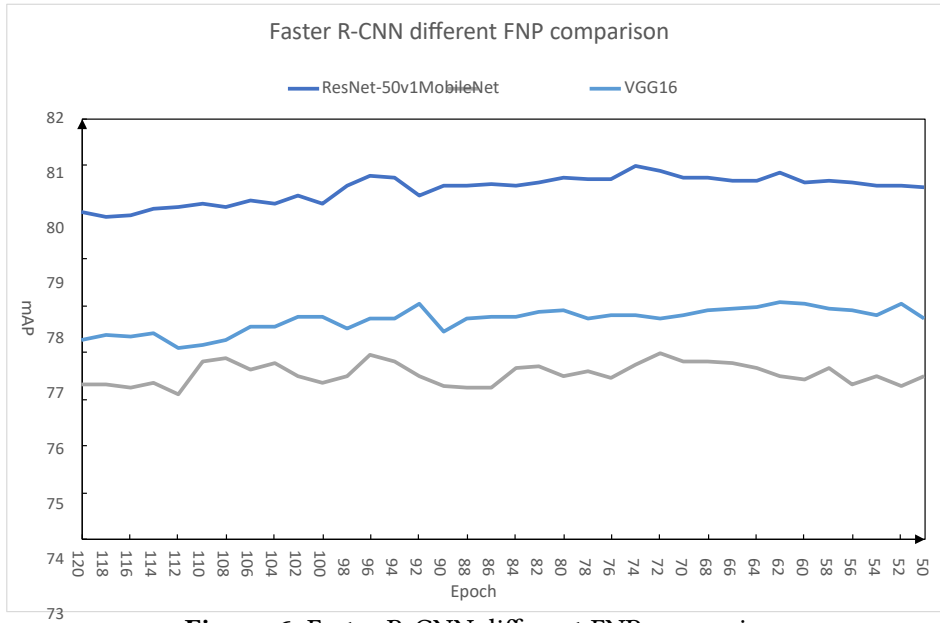
Figure 6: Faster R-CNN different FNP comparison.

The real-world testing were done using 100 custom annotated frames from sport video footage with 664 logos in it. The model detected 523 logos bounding boxes – 78% of all logos that were in testing data. The results presented in the confusion matrix Figure 7.

Figure 7: Testing results confusion matrix.

5. Conclusions

In this study explored a deep learning algorithms to evaluate brand visibility in live sports videos, presenting a novel approach to address the challenges of manual annotation and subjective analysis. Through the utilization of advanced object detection models like YOLO and Faster R-CNN, the final model demonstrated the capability of automated methods to accurately detect and localize brand logos within the dynamic context of sport videos. Even though the final model demonstrated the accuracy of detecting eight out of ten logos in real – world video, this finding emphasizes the significant potential of automated solution in overcoming the limitations associated with manual annotation, offering more objective and more efficient evaluation of brand positioning. Looking forward, the continued development and refinement of deep learning methodologies, coupled with advancements in real-time monitoring capabilities, hold promise for further enhancing the accuracy and effectiveness of brand visibility evaluation in live sports videos. Additionally, the availability of larger and more diverse annotated datasets will be necessary in improving model performance and generalization.

6. Acknowledgment

I want to express my gratitude for my supervisor prof. Agnė Paulauskaitė – Tarasevičienė for encouragement and valuable insights.

7. References

[1]   L. Turley and J. Richard Shannon, "The impact and effectiveness of advertisements in a sports arena," *Journal of services marketing.*

[2]   S. Shaz, Experiential Marketing. A practical guide to interactive brand experiences, London and Philadelphia: Kogan Page.

[3]   G. Raghav and G. Aradhana, "Impact of Elements of Ad's on Sports Fan Attitude during a Live Sporting Event," *Scholarly Journaly,* no. 24, pp. 867 - 876, 2022.

[4]   R. Arora, A. Chawla and V. Sachdeva, "An analytical study of consumer awareness," *International Journal of Advanced Research in Management and Socal Sciences,* vol. 13, no. 8, pp. 137 - 153.

[5]   L. Yuan, L. Xiaoqing, Z. Chengcui, W. Yongtao and T. Zhi, "Mutual enhancement for detection of multiple logos in sports videos," in *IEEE International Conference on Computer Vision*, 2017.

[6]   R. Boia, C. Florea and L. Florea, "Elliptical ASIFT agglomeration in class prototype for logo detection," in *Proceedings of the British Machine Vision Conference*, 2015.

[7]   C. Wan, Z. Zhao, X. Guo and A. Cai, "Tree-basedShapeDescriptorforscalablelogodetection," in *Visual Communications and Image Processing*, 2013.

[8]   H. Sujuan, L. Jiacheng, M. Weiqing, H. Y. Z. Qiang, Z. Yuanjie and J. Shuqiang, "Deep Learning for Logo Detection: A Survey," *ACM Trans. Multimedia Comput. Commun,* vol. 20, no. 72, 2023.

[9]   Y. Kangning, H. Shaoqi, L. Ye, L. Chao and Y. Guangqiang, "A real-time vehicle logo detection method based on improved YOLOv2," in *Proceedings of the International Conference on Wireless Algorithms, Systems, and Applications.*

[10]   J. Redmon, S. Kumar Divvala, R. B. Girshick and A. Farhadi, "You only look once: Unified, real- time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[11]   C. Egget, D. Zecha, S. Brehm and R. Lienhart, "Improving Small Object Proposals for Company Logo Detection," in *ACM on International Conference on Multimedia Retrieval*, New York, 2017.

[12]   C. Egget, S. Brehm, A. Winschel, D. Zecha and R. Lienhart, "A closer look: Small object detection in faster R-CNN," in *IEEE International Conference on Multimedia and Expo*, Hong Kong, 2017.

[13]   A. Alsheikhy and Y. Sald, "Logo Recognition with the Use of Deep Convolutional Neural Networks," in *Engineering, Technology and Applied Science Research*, 2020.

[14]   C. Wilms, R. Heid, M. Araf Sadeghi, A. Ribbrock and S. Frintrop, "Which airline is this? Airline logo detection in real-world weather conditions," in *Proceedings of the 25th International Conference on Pattern Recognition*, 2021.

[15]   F. Sultana, S. Abu and D. Paramartha, "A Review of Object Detection Models based on Convolutional Neural Network," *ntelligent computing: image processing based applications,* pp. 1-16, 2020.

[16]   X. Long and ..., "PP-YOLO: An effective and efficient implementation of object detector.," *rXiv preprint arXiv:2007.12099,* 2020.

[17]   A. Joly and O. Buisson, "Logo retrieval with a contrario visual query expansion," in *roceedings of the 17th ACM International Conference on Multimedia.*

[18]   S. Romberg, L. Garcia Pueyo, R. Lienhart and V. Z. Roelof, "Scalable logo recognition in real - world images," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval.*

[19]   A. Kuznetsov and A. V.Savchenko, "A new sport teams logo dataset for detection tasks," in *Proceedings of the International Conference on Computer Vision and Graphics.*

[20]   H. Su, X. Zhu and S. Gong, "Open logo detection challenge," in *Proceedings of the British Machine Vision Conference.*

[21]   H. Qiang, M. Weiqing, W. Jing, H. Sujuan, Y. Zheng and J. Shuqiang, "FoodLogoDet-1500: A dataset for large-scale food logo detection via multi-scale feature decoupling network," in *Proceedings of the 29th ACM International Conference on Multimedia.*

[22]   W. Jing, M. Weiqing, H. Sujuan, M. Shengnan, Z. Yuanjie and J. Shuqiang, "LogoDet-3K: Alarge- scale image dataset for logo detection," in *IEEE Transactions on multimedia.*