A novel approach to machine learning for object detection and recognition

Bh.Sai Venkata Ganesh [1*], N. Siva Kumar[2], Bh.Sai Venkata Ganesh[1], N. Siva Kumar[2]

[1] Aditya University, Surampalem, Andhra Pradesh
[2] BVC Group of Institutions, Amalapuram., Andhra Pradesh
*svganesh44@gmail.com

Abstract: Artificial intelligence (AI) in computer science is a branch that focuses on creating intelligent systems or robots capable of imitating human behavior and reactions. Artificial intelligence is a branch of computer science. The impressive ability of individuals to easily identify and differentiate items using their visual senses is astonishing. However, robots have many substantial obstacles in detecting and identifying objects. Neural networks are a suggested solution from the field of computer science to address this problem. Additionally, it is sometimes known as "Artificial Neural Networks." These words are all used interchangeably. Neural networks are an example of artificial intelligence that operates without the usage of symbols. The purpose of these computer models is to simulate the operations of the human brain to assist in the identification and classification of various objects. Conversely, object detection and identification is a field that undergoes extensive investigation. This research focuses mostly on dynamic things, which are objects in motion. The system is meant to perform object detection and static object identification to achieve its specified purposes. Our solution replaces the basic classifier from the previous system with a more advanced one, resulting in an increased accuracy rate. The Object Detection and Tracking System will use the renowned deep learning network Faster Regional Convolution Neural Network (Faster R-CNN) for object detection. This implies that the system will have the capability to identify items. Moreover, the traditional object tracking mechanism will be used in this specific project. By using closed-circuit television cameras in tunnels, it will enable the automated detection and recording of any unforeseen events. The Object Detection and Tracking System used deep learning to carry out its functions. This model was trained on a dataset consisting of tunnel event photos captured by photography. The model attained mean accuracy values of 0.8479, 0.7161, and 0.9085 for detecting fire target items. The values were acquired for detecting autos, persons, and firing targets. The model generated these numbers via its computations. The Tunnel CCTV Accident Detection System, based on ODTS, was tested by analyzing four accident videos, each containing a separate accident, using a trained deep learning model.

Keywords: RCNN algorithm, YOLO, Optimization, Neural Networks, Computer Vision, Image Processing.

K. R. Madhavi et al. (eds.), *Proceedings of the International Conference on Computational Innovations and Emerging Trends (ICCIET 2024)*, Advances in Computer Science Research 112,
https://doi.org/10.2991/978-94-6463-471-6_87

1 Introduction

The process of identifying occurrences or things that differ from the norm by exhibiting qualities that are irregular, unexpected, unpredictable, and unusual is referred to as anomaly identification. Within the context of this procedure, the term "anomaly detection" refers to the mechanisms that are involved. Keeping an eye out for the abnormality's features is one approach that might be used to achieve this objective. This technique is also referred to as a weapon identification system, and it is frequently used within a certain segment of the community. Another term that may be used to describe this occurrence is anomaly detection, which is another name for it. This is another word that can be used about it. These instances or objects are regarded to be non-standard occurrences or atypical things since they deviate significantly from the established patterns or items that are already included in a collection. This is the reason why they are deemed to have this classification. They are thought to be out of the ordinary because of this reason. As a result of this, it is generally accepted that they are the correct explanation. Because they are not a part of the collection that is being considered, this is the reason why this is the case. The term "anomaly" refers to a pattern that serves the purpose of distinguishing itself from the collection of regular patterns that are often recognized. The discipline of pattern analysis makes use of the term "anomaly" in its vocabulary. In every instance in which the pattern is being addressed, the word "anomaly" is the one that is used to describe it. It is vital to consider the features of the phenomenon that is being studied to ascertain whether there are any anomalies that could arise in the phenomenon that is the focus of the investigation. This is because the outcome of the investigation will determine whether there are any abnormalities. Object detection is a computational approach that detects and categorizes numerous instances of things while concurrently gathering crucial information from these items. This method is also known as "object detection." The completion of these tasks is accomplished using learning algorithms or models using this technique. In the context of this discussion, the word "object detection" refers to a method that is founded on the implementation of computing. In the process of object detection, there are two distinct components that may be separated from one another: the first is the identification of objects, and the second is the classification of those items. The process of object detection may be broken down into these two independent components. A crucial component, which is the precise identification and categorization of the many kinds of guns, is included in the implementation that is advised. Within the implementation, this component is among the most crucial parts. As a result of the fact that an incorrect warning could bring about unfavorable outcomes, it is of the utmost importance to make certain that the information given is accurate and comprehensive. Finding a method that strikes a harmonic balance between the speed with which the technique is performed and the accuracy with which it is administered is that which is required to get a method selection that can be regarded acceptable. When it comes to the process of acquiring a technique selection that can be regarded as acceptable, this is a vital phase. At the same time as the technique for input is being carried out, the process of extracting frames from the video is also being carried out simultaneously. Following the completion of the frame differencing approach and the acquisition of a bounding box, the second stage, which is the identification of an object, involves carrying out the process. Following the conclusion of the stage that came before it, this stage will now begin. The subsequent phase is brought about by this stage, which comes after the differentiation of the frames occurred.

2 Literature Review

Now, academics make use of fundamental datasets for a variety of purposes. Among these procedures are the identification and tracking of pedestrians, as well as the evaluation of behavior that may be considered suspicious. There are 2,300 distinct pedestrians included in the dataset that was produced by Caltech, and there are also 350,000 bounding boxes that have been labeled to specifically represent these people. A camera that was installed on a vehicle was used to record video footage of roadways in urban areas, which was then combined to form the dataset. The dataset itself was created with the help of the film. One of the most well-known collections of pedestrian data is the MIT dataset, which is comprised of photographs of people walking that have been taken with great care. As far as estimates are concerned, there are around 709 distinct kinds of pedestrians taken into consideration. There are only a few positions that can be captured in pictures shot on city streets, and this is true regardless of whether the images are taken from the front or the rear of the subject. The purpose of this dataset is to use cameras that are installed on automobiles in a metropolitan region during daylight hours to monitor the movement of individuals on public highways.

3 Methodology

Joseph Redmon originally proposed the idea that would later develop into the YOLOv1 object detection paradigm in 2016. Although it is not essential for the YOLOv1 detection model to operate well, it is highly recommended to do the area suggestion extraction procedure. There is no need to eliminate this method in any way. When simplified, the detection model is just a standard CNN network structure. There is a chance of seeing the whole model at this specific location. Utilizing the whole graph as the network input and directly outputting the location and category of the bounding box are two essential notions that underpin this technique. Both notions are crucial for using this technology. This phrase combines both notions into one remark. An example of how this approach may be implemented is shown by executing any of these concepts. Each image is first divided into an S*S grid. Each cell inside the grid is responsible for predicting a B-bounding box and providing confidence ratings for these boxes. This technique is continued until all photos are divided into S*S grids. The method is done once every picture has been evaluated. This procedure is repeated until that juncture. One may argue that each cell calculates an educated approximation of the total B*(4+1) values. This is an alternative. Another potential interpretation of these data remains. In other words, this is another way of conveying the same idea. The detection speed may approach 45 frames per second on a single TitanX, enabling real-time detection. This is another benefit of TitanX. There is a possibility of this happening. Furthermore, TitanX offers some significant benefits. Although YOLO generates fewer errors in the background, it has a limited ability to recognize objects in densely packed scenes. Even if it has a lower overall mistake rate, this still applies. Nevertheless, this occurs despite a reduced total of mistakes. YOLO handles information sequentially compared to other approaches.

.

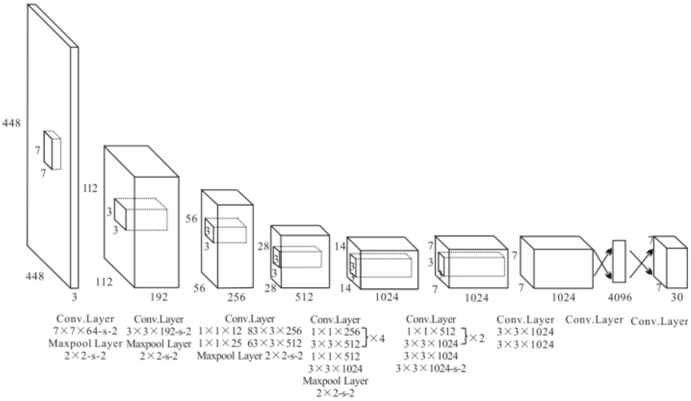

Figure 1. YOLOv1 architecture

3.1 SSD

In 2016, Liu proposed solid-state drives (SSD) as a potential solution to an emerging issue. The Faster R-CNN detection model introduced the anchor box concept, which served as inspiration for this discovery. This was crucial for attaining the development's goals. Moreover, the YOLO method incorporates the idea of regression into the model with the goal of enhancing its inclusivity. Furthermore, this issue was previously brought up throughout the discussion. The SSD method suggests using both low-level and high-level feature maps for detection to enhance the efficiency of detecting objects at various scales. This improves the overall efficiency of the detection process. This step may be taken to improve performance. This activity tries to achieve peak performance by pursuing the highest level of performance. The VGG architecture serves as the basis of the layer hierarchy, where the convolutional layers take the place of the two fully connected layers that come after them in the hierarchy. This is because the VGG architecture serves as the foundation. The solid-state drive (SSD) utilizes an anchor mechanism to provide the RPN network with reliable functioning. An SSD achieved a mean average precision (mAP) of 74.3% on the VOC2007 test while running at 59 frames per second on an Nvidia Titan X. This is a notable accomplishment. This achievement has enormous importance. SSD's classification performance is inadequate for extremely tiny targets because the feature maps at various scales are independent, resulting in the detection of the same item by boxes of different sizes concurrently. On the other hand, the SSD classification result is exceptional for very large targets. The attribute mappings are mutually exclusive, resulting in this conclusion. Feature maps of different scales are believed to be independent of one another, which helps to understand the situation. The SSD classification performance is insufficient for extremely tiny targets, which harms the existing damage.

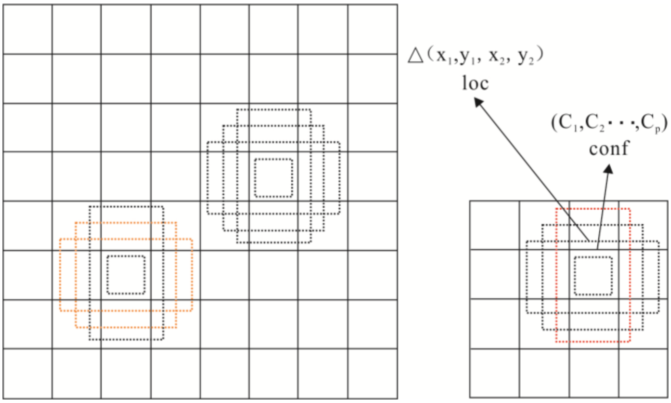

Figure 2. SSD architecture

4 Results and Discussions

The phrase "artificial intelligence" was first used in 1956 for study and discourse. In 2012, artificial intelligence reached its peak for the first time. This event took place in 2012. Several causes contribute to this phenomenon, including the increasing amounts of data, advancements in processing capabilities, and the development of machine learning algorithms tailored for this task. Research has shown a significant correlation between the advancement of new detection technologies and the increase in data volume. A profound connection is formed between the two as a result. This circumstance exists because they have a strong relationship. This dataset is essential for conducting performance testing and algorithm evaluations, as it serves as a stimulus for developing detection algorithms and is crucial for their assessment. This is the reason behind the present situation. Table 1 outlines the bulk of datasets accessible to the public and offers information on their features.

Table I. Public Data Set And Its Parameters

Dataset	**Amount**	**Sort**	**Size/Pixel**	**Year**
Caltech106	8547	101	300×200	2015
PASCAL VOC 2010	1063	20	375×500	2016
PASCAL VOC 2016	18623	20	470×380	2017
Very small Images	79.5 million	56899	32×32	2017
Scenario17	5962	15	256×256	2018
Kaltechi56	32054	256	300×200	2019
ImageNet	16248568	21841	500×400	2020
RUN	235485	908	500×300	2021
MS COCO	469754	91	640×480	2022

Place	Greater than 12 million	434	256×256	2023

5.2 Comparative analysis of the efficacy of several strategies

Table 2: A comparison of the object detection solutions which are currently available.

Method	**Backbone**	**Size/Pixel**	**Test**	**mAP/%**	**fps**
YOLOv1	VGG17	548×548	VOC 2008	71.4	44
SSD	VGG17	400×400	VOC 2008	80.2	48
YOLOv2	Darknet-21	644×644	VOC 2008	82.3	39
YOLOv3	Darknet-61	708×508	MS COCO	40	52
YOLOv4	CSP Darknet-52	708×508	MS COCO	48.7	67.8
R-CNN	VGG16	1100×700	VOC2008	71	0.6
SPP-Net	ZF-6	1100×700	VOC2008	60.2	-
Fast R-CNN	VGG13	1100×700	VOC2008	77.9	10
Faster R-CNN	ResNet-109	1100×700	VOC2008	80.02	5

5 Conclusion

In recent years, there has been a notable rise in the focus on object detection. This is primarily because it is one of the most basic and challenging subjects in computer vision. This topic is demanding since it is one of the hardest. Although deep learning has been used in several domains using detection algorithms, it still encounters substantial barriers that need more investigation. Some of these concerns include:Engage in activities that will lessen the degree to which you are dependent on data.To accomplish the identification of minute things in a manner that is both effective and efficient.The carrying out of processes for the detection of items across a variety of distinct categories.

6 References

[1] Ahmed M., Jahangir M., Afzal H. (2015) “Using Crowd-source based features from social media and Conventional features to predict the movies popularity”, IEEE International Conference on Smart Cities, Social Community and Sustained Community, China, pp. 273–278.

[2] Bergmann P., Meinhardt T., Taixe L. (2019) “Tracking without bells and whistles”, IEEE International Conference ICCV, Seoul, Korea, pp. 1-16.

[3] Dollar P., Wojek C., Schiele B., and Perona P. (2012) “Pedestrian Detection: An Evaluation of the State of the Art”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 34 (4): 743-761.

[4] Samsi S., Weiss M., Bestor D., Li D., Jones M., Reuther A., Edelman D., Arcand W. and Byun C. (2021) “The MIT Supercloud Dataset”, International Cornell Journal of Distributed, Parallel, and Cluster Computing, 2108 (02037): 1-10.

[5] Silberstein S., Levi D., Kogan V. and Gazit R. (2014) “Vision-based pedestrian detection for rear-view cameras”, IEEE Intelligent Vehicles Symposium, pp. 853-860, Dearborn, MI, USA.

[6] Alom M. and Taha T. (2017) “Robust multi-view pedestrian tracking using neural networks”, IEEE National Conference on Aerospace and Electronics, pp. 17-22, Dayton, OH, USA.

[7] Zhang X., Park S., Beeler T., Bradley D., Tang S., Hilliges O. (2020) “ETH Gaze: A Large-Scale Dataset for Gaze Estimation Under Extreme Head Pose and Gaze Variation” European Conference on Computer Vision (ECCV), Springer. Lecture Notes in Computer Science, pp. 1-10.

[8] Wojek C., Walk S., Schiele B., (2009) “Multiresolution model for Object Detection”, IEEE Computer Vision and Pattern Recognition (CVPR), June 20-25, 2009, Miami, Florida, USA.

[9] Nguyen T., Soo K., (2013). “Fast Pedestrian Detection Using Histogram of Oriented Gradients and Principal Components Analysis”. International Journal of Contents, 2013 (1): 1-20.

[10] Everingham, M., Van L., Williams, C. (2010) “The Pascal Visual Object Classes (VOC) Challenge”, International Journal of Computer Vision, Springer, 88 (1): 303–338.

[11] Lin T. (2014), “Microsoft COCO: Common Objects in Context. ECCV 2014”, Lecture Notes in Computer Science, Springer, 8693 (1): 740- 755

[12] Nicolai W., Bewley A., Dietrich P. (2017) “Simple online and real-time tracking with a deep association metric”, IEEE International Conference on Image Processing (ICIP), pp. 3645–3649.

[13] Jifeng D., Yi L., Kaiming H., Sun J., (2016) “R-FCN: Object detection via region-based fully convolutional networks”, IEEE Computer Vision and Pattern Recognition (CVPR), pp. 1-11.

[14] Everingham, M., Eslami, S., V. G., Williams C., Winn J. (2015) “The PASCAL VOC Challenge: A Retrospective”, International Journal of Computer Vision (IJCV), Springer, 11 (1): 98-136.

[15] Kaiming H., Georgia G., Dollar P. and Girshick R. (2020) “Mask R-CNN, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 42 (2): pp. 386-397.

[16] Raju, S. Viswanadha, A. Vinaya Babu, G. V. S. Raju, and K. R. Madhavi. "W-Period Technique for Parallel String Matching." IJCSNS 7, no. 9 (2007): 162..

[17] Felzenszwalb P., Girshick R., McAllester D. and Ramanan D. (2010) “Object detection with discriminatively trained part- based models”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 32 (9): 1627–1645.

[18] Dalal N. and Triggs B. (2005) “Histograms of oriented gradients for human detection”, IEEE Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, pp. 886–893.

[19] Muhammad N., Hussain M., Muhammad G., Bebis G., (2011), “Copy-move forgery detection using dyadic wavelet transform,” International Conference on Computer Graphics, Singapore, pp. 103–108.

[20] Muhammad G., Hossain M., Kumar N., (2021) “EEG-based pathology detection for home health monitoring”, IEEE Journal on Selected Areas in Community, 39 (2): 603–610.

[21] Muhammad G., Alhamid M., and Long X., (2019) “Computing and processing on the edge: Smart pathology detection for connected healthcare”, IEEE Network, 33 (1): 44–49.

[22] Girshick R., Donahue J., Darrell T., and Malik J., (2014), “Rich feature hierarchies for accurate object detection and semantic segmentation” IEEE Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, pp. 580–587.

[23] He K., Zhang X., Ren S., and Sun J., (2015) “Spatial pyramid pooling in deep convolutional networks for visual recognition”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 37 (9): 1904–1916.

[24] Girshick R., (2015) “Fast R-CNN”, IEEE International Conference on Computer Vision, Santiago, Chile, pp. 1440–1448.

[25] Ren S., He K., Girshick R. and Sun J., (2016) “Faster R-CNN: Towards real-time object detection with region proposal networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 39 (6): 1137–1149.