iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** VII **Month of publication:** July 2023

DOI: https://doi.org/10.22214/ijraset.2023.54966

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue VII Jul 2023- Available at www.ijraset.com

Various Approaches to Object Detection using Deep Learning

Somil Doshi[1], Krishna Desai[2], Kush Mehta[3]
[1, 2]*Information Technology, Thakur College of Engineering and Technology,*
[3]*Computer Engineering, K.J. Somaiya Institute of Technology*

Abstract: ***In order to detect objects in actual scenarios, this research study examines the practical use of deep learning techniques, with a focus on convolutional neural networks (CNNs). The importance of object categorization as the fundamental building block for object detection is emphasized, and several learning strategies are investigated in order to extract features from challenging multidimensional data. In order to get accurate and effective object detection findings, the study studies various deep learning architectures, including R-CNN, Fast R-CNN, Faster R-CNN, and YOLO. In addition, this paper demonstrates the diverse uses of deep learning-based object detection, including its influence on automated vehicles, medical technology, and industrial automation, where it is changing several sectors of the economy. The analysis also foresees the future developments of deep learning models, predicting their paradigm-shifting effects across a variety of areas and their capacity to successfully address practical difficulties.***

Keywords: Object Detection, Convolutional Neural Network (CNN), Faster R-CNN, Deep Expectation (DEX), YOLO, Deep Neural Network (DNN)

I. INTRODUCTION

For a scenario, there are usually a number of things available. Each thing possesses its own collection of qualities. The traits are distinctive and include information about the object. The Object classification method is used to categorize it. Object classification serves as the foundation for object detection since each object in the scene is given a label. To find features, algorithms that can learn from the dataset might be employed. Humans can learn by employing the information they have learned, and learning and knowledge representation go hand in hand. People are able to interpret and comprehend the data. There are supervised, unsupervised, and semi-supervised learning techniques. Fascinating facts are delivered by data, some of which may be multidimensional. Deep learning or machine learning algorithms can use the features as input. [3] The use of generic characteristics for categorization is made possible by deep learning techniques. To learn the features at a context level, stochastic models might be utilized. Iterative testing and training can be used to cut down on losses. The type of scene is important for object identification.

The objects can also be better learned based on the related objects in the scene. The object's position, colour, mobility, and eyesight are all problematic. Algorithms for motion suppression can regulate the motion. The study provides a comprehensive examination of using deep learning techniques to object detection.

Smart surveillance systems employ artificial intelligence (AI) to automate the surveillance of both industrial and residential structures. To recognize and identify the presence of certain objects in the field of view of the CCTV camera, most intelligent surveillance systems use deep learning and machine learning. Due to the real-time nature of smart systems, the object detection approach must abide by a number of restrictions. For instance, a smart surveillance system's efficacy depends on minimizing the time required by the algorithm to make predictions (the inference time). [6] Deep learning is now a useful, superior surveillance technology since experiments involving item identification and recognition have shown that deep learning algorithms perform better than human accuracy. In terms of security, surveillance is essential. Despite evidence to the contrary, surveillance methods and procedures still need to be improved before they can be successfully implemented into smart settings. The transportation sector, which includes airports, railway stations, roadways, labs, and hazardous chemical containment facilities, is one of many that greatly benefits from smart space equipment. There is now sufficient data to demonstrate that machines frequently outperform humans in the realm of surveillance, thus we must focus on applying deep learning for object detection.

The fact that the cost of monitoring will go down as human beings are eventually replaced by robots is another benefit that stimulates more in-depth research in the field. [1] Until deep learning is accurate enough to do so affordably and economically, it won't replace current computer vision and human surveillance techniques. It takes a lot of datasets to train the algorithm because there is so much previously shot content that is now accessible.

©IJRASET: All Rights are Reserved

Many models have been trained and used successfully to find and classify items in images and subsequently movies. In many cases, these models employ an architecture that involves the input image using a CNN before sending it to a regional proposal network for classification. A region proposal network examines the input image to identify which subframe most likely contains the object of interest by differentiating between the background and the object. The result is what is important. The workflow of these systems, which was previously described in more general terms, is shown in Fig. 1. This study analyses the many models that are now available based on the published values of key performance criteria.[5]

II. LITERATURE SURVEY

Initially created in the 1990s, convolutional neural networks use supervised learning to classify tagged images. In contrast to traditional neural networks, which can only process input vectors, CNNs can handle inputs of greater dimensions. A CNN can therefore be simply applied to a three-dimensional input image as a result. Shared parameters and sparse connections are the main advantages of CNNs. [2] CNNs typically consist of a convolutional layer with filters that combine with the input image. A pooling layer receives the output from this system after a predetermined number of these layers. As a one-dimensional vector, this can be further flattened and given to neural networks.

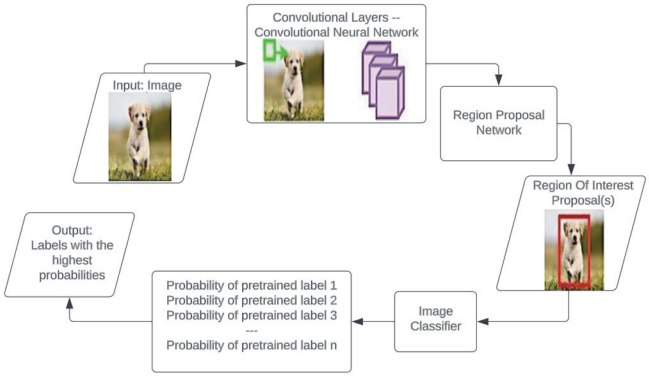


Fig 1: Flowchart for region localization & object detection models

The fundamental concept of CNNs has several crucial applications, with surveillance being one among them. Given that existing models are becoming more accurate and that deep learning can completely automate monitoring, this is a very real scenario. The dangers that must be identified in a smart environment will commonly fall into a common category (such as fire, smoke, an unknown individual, etc.), making it easier to train a deep learning model with high accuracy. Because the cost of implementing these deep learning models has been steadily declining, artificially intelligent surveillance systems can be deployed widely. A brief summary of the significant developments in the field over the preceding ten years serves to illustrate this. ImageNet organizes the ILSVRC, an annual worldwide contest.[14]

Finding out more about various object recognition and image classification methods by looking at the competition entries is beneficial. Comparing entries involves using the top-5 classification and localization error rates. Regarding its overall categorization technique and architectural design, each entry is unique. Different models employ various layer types, numbers of each layer, orderings, and hyperparameters like filter size and epochs. The most notable models are listed chronologically below:

AlexNet SuperVision performed better in 2012 than the industry average top-5 error rate, which was 15.3%. In the same year, a 6% error rate for the human classifier was recorded. Be aware that the theoretical premise of this study is that test error rate = 1 - test accuracy holds true. With an error of 11.7% in 2013, the ZFNet significantly lowered this number. In 2014, GoogLeNet followed suit with an error of 6.6%. An important aspect of the 2014 competition was the discovery that the winning model in the classification competition was not the best model for localization. Google's convolutional neural network was able to accurately classify dog breeds, while Karpathy, a human classifier, only managed an error of 5.1%. In February and December of 2015, Microsoft developed two models that reduced the error rate to 4.94% and 3.5%, respectively, effectively outperforming human accuracy for the first time ever. Trimps-Soushen developed and specified his model in Chinese with a 2.99% error rate. The SENet now has the lowest error of any model, at 2.251%. [13] These models are now particularly useful in the fields of surveillance and security measures since they are so accurate and persistent in comparison to human efforts, and because well-trained algorithms are frequently unbiased. A comparison of the error rates attained by deep learning models and people is shown in Figure 2.

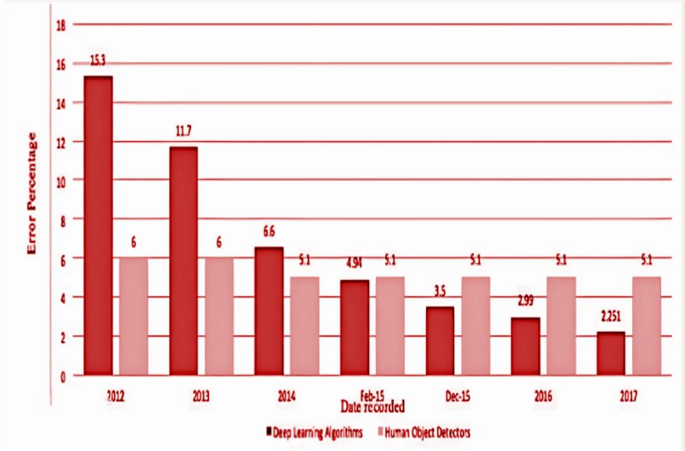


Fig 2: Deep learning algorithms' error rate in object classification compared to human error [14]

R-CNNs have improved their ability to recognize fine-grained classifications. This is done by utilizing cutting-edge localization methods. This paper shows that bottom-up region concepts outperform end-to-end review. According to reports, the scalable detection algorithm's ensemble model approach increased the map by 30%. The author of this article presents a method for properly predicting 3D bounding boxes in photographs with severe distortion and discusses recent developments in the field of 3D object detection. More recently, introduces an algorithm that can identify salient parts in the video and offers data augmentation methodologies and the use of salient area information to prevent overfitting, which is a common problem with powerful neural networks combined with huge datasets.

III. RESEARCH GAP

Despite the significant advancements and prominent successes in deep learning-based object detection, there are still a number of research gaps that offer opportunities for further investigation and advancement. Some of the major research gaps in object detection using deep learning include the ones listed below:

1) *Small Object Detection:* Deep learning algorithms commonly struggle to identify small things in pictures because they lack spatial information. The development of techniques and architectures that can precisely detect and recognize little items will significantly enhance the performance of object identification.[2]
2) *Occlusion Handling:* Deep learning models commonly struggle when objects are partially hidden by other objects or dense backgrounds. Object identification systems must be strengthened to handle occlusions and locate objects precisely even in challenging conditions.
3) *Generalization to Unseen Domains:* Many deep learning models for object detection may struggle to generalize to new or domain-shifted data, even having been trained on big datasets. Addressing the issue of domain adaptation and transfer learning is crucial for real-world applications where test data and training data could differ.[3]
4) *Effective Real-Time Detection:* While many deep learning models are computationally expensive, preventing their broad application in resource-constrained situations, some of them are capable of real-time detection. It would be beneficial to look into more efficient architectures and model compression techniques in order to enable real-time item recognition on edge devices.[6]
5) *Few-Shot and Zero-Shot Object Detection:* Typical deep learning models require a significant amount of labelled training data. If zero-shot and few-shot learning approaches were examined for object detection, models might be able to recognize items without any prior training or even learn from little quantities of labelled data.[5]
6) *Resistance to Adversarial Attacks:* It is common knowledge that Adversarial Attacks can lead to misclassification in deep learning models for object detection. More trustworthy and adversarial-resistant models must be created in order for object detection systems to be deployed in security-critical applications.
7) *Explainability and Interpretability:* Because many deep learning models are thought of as "black boxes," it could be challenging to understand how they get to their conclusions. By looking for ways to make the predictions and interpretations of the models simpler to understand, object detection algorithms' credibility and acceptance can be increased.[9]

8) *Multi-Object Tracking:* Unlike object detection, which focuses on finding objects in specific frames, multi-object tracking aims to maintain object IDs across frames in movies or sequential data. The gap between object identification and multi-object tracking needs to be filled in order to fully comprehend a scene.
9) *Handling Unbalanced Data:* If specific object classes are significantly underrepresented, unbalanced datasets may result in biased models and subpar performance on minority classes. Investigating strategies to lessen class disparity and improve the detection of rare objects is crucial for practical applications.[10]
10) *Domain-Specific Object Detection:* Depending on the domain, object detection can provide specific challenges and requirements. It is possible to develop more specialized and ideal solutions for a variety of applications, such as industrial automation, autonomous vehicles, and medical imaging, by looking at domain-specific designs and methods.

IV. ARCHITECTURE

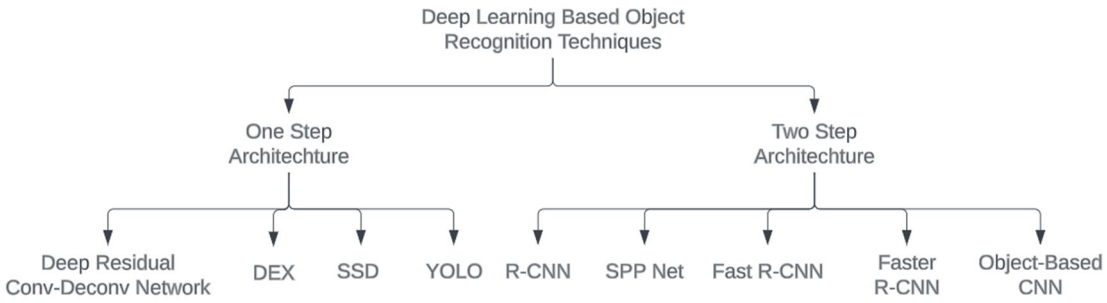


Fig 3 : Overview of object detection techniques [8]

A. One Step Architecture

Szegedy et al. proposed a simple bounding box interference and binary mask-based deep neural network (DNN) technique for object abstraction. However, it is challenging to efficiently extract overlapping objects using this approach. In response, Pinheiro et al. created a CNN model with two branches. The first branch produces masks, whereas the second branch predicts the likelihood that an object would be present in a specific location. Erhan et al. developed a multibox approach based on regression to provide region proposals while Yoo et al. recommended using AttentionNet, a CNN architecture, for object recognition using a classification strategy.[8]

In the realm of age estimation, the Deep Expectation (DEX) technique was introduced. It does not rely on facial landmarks and instead uses the IMDB-WIKI database of face pictures with age and gender characteristics. The method takes use of a CNN with a trained VGG-16 architecture on ImageNet. Faces are aligned via angle changes and cropping as opposed to facial landmarks, which have a high failure risk. The aligned face is then processed by the VGG-16 CNN to determine age. Ranges are used to express the age value in order to categorize the age forecast. To improve CNN, the IMDB-WIKI dataset is employed.

In the context of hyperspectral image analysis, a novel neural network design known as the Deep Residual Conv-Deconv Network is proposed for unsupervised learning of spectral spatial features. The encoder (convolutional subnetwork) transforms the 3D hyperspectral data into a lower dimensional space, while the decoder (deconvolutional network) reconstructs the data to its original state. This is the basis of the fully convolutional deconvolutional network. The network uses unique un-pooling techniques and residual learning to improve performance. The convolution subnetwork consists of a number of convolutional blocks with increasing channel dimensions. The deconvolutional network reconstructs the input data using the deep features. The learning latency and edge data depletion during decoding must be addressed by network optimization.

The YOLO (You Only Look Once) framework developed by Redmon et al. generates predictions on the confidence of bounding boxes and other categories. It divides the image into grids and makes educated predictions about the elements that make up its core in each grid cell using bounding boxes and confidence scores. YOLO generates coarse features and has problems handling clusters of little objects as a result of numerous down sampling techniques.

Liu et al. developed the Single Shot Multibox Detector (SSD) methodology to overcome some of the limitations of the YOLO approach. Instead of using fixed grids to discretise the output space of bounding boxes, SSD use anchor boxes with various aspect ratios and sizes. Numerous feature maps with varied resolutions can forecast the offset of default boxes with different scales and aspect ratios. The SSD design employs feature layers connected to the VGG16 network for prediction and a weighted sum of localization and confidence losses for training.

The final item detection is achieved using non-maximum suppression (NMS) on multiscale enhanced bounding boxes.[8] The aforementioned techniques offer fresh approaches to age estimation, object detection, and hyperspectral picture analysis. While some methods make use of CNNs to address specific issues, others use state-of-the-art network topologies to improve feature learning or address issues with down sampling. These innovations advance computer vision as a whole and pave the way for more accurate and successful object detection and recognition across a range of applications.

B. Two Step Architecture

Ross Girshick developed R-CNN in 2014 as a method for improving candidate bounding boxes' quality and extracting high-level features from deep architecture. Results produced by R-CNN on the PASCAL VOC 2012 dataset were superior to those of older techniques. It consists of two stages: region proposal creation and CNN-based deep feature extraction. Approximately 2K region concepts are generated during the region proposal stage using the selective search technique. Choose search uses bottom-up grouping and saliency cues to quickly construct exact bounding boxes of any size. In deep feature extraction, cropped region proposals are used to extract solid 4096-dimensional features using a deep CNN. [16] Category-specific linear SVMs are used to identify the positive and negative regions, and bounding box regression is then used to further refine the regions. To produce the final bounding boxes for object localisation, these regions are filtered using greedy non-maximum suppression.

R-CNN had several benefits over traditional approaches, but it also had significant disadvantages. Due to the need for a fixed-size input image during testing, the CNN had to be laboriously recalculated. The multi-stage R-CNN training process required a lot of time and storage. [19] Due to the high amount of region proposal duplication, the process took a long time, and information loss and distortion happened when the region proposals were distorted or cropped to the required size, lowering the recognition accuracy. He and his collaborators created SPPnet, which uses the 5th convolutional layer (conv 5) to express random-sized region proposals as fixed-size feature vectors. A spatial pyramid pooling layer (SPP layer), which was added after the final convolutional layer, produced feature vectors that were more useful. However, SPPnet continues to encounter storage space constraints and an accuracy reduction in deep networks due to the inability to update the convolutional layer before the SPP layer.

Girshick proposed Fast R-CNN as a remedy for these problems. The convolutional layer in this technique uses the full image to analyse and generate feature maps. Each region proposal is processed via a ROI pooling layer to extract fixed-length feature vectors. These feature vectors are then processed through many fully connected layers to produce category probabilities and bounding box positions. Truncated singular value decomposition (SVD) is used in the fc layers to speed up the pipeline. Following that, the RPN was used to predict item placements in other object detection networks, leading to a quicker R-CNN that ran for shorter periods of time.[8]

The semantic gap of complex patterns in the context of disaster response and urban planning utilizing high-resolution pictures can be effectively closed using deep learning. However, object boundaries are hard to capture using deep learning techniques. A method is suggested to increase the precision of high-resolution picture categorization by combining deep feature learning methods with object-based classification. The two steps in this strategy are deep feature extraction using CNN and object-based classification using these features. The effectiveness and precision of the object recognition techniques R-CNN, SPPnet, Fast R-CNN, and Faster R-CNN have continually improved. The effectiveness of feature extraction and recognition has significantly increased because of CNNs and deep learning. [9] However, problems like the need for fixed-size input and processing costs persisted. The proposed combination of deep feature learning and object-based categorization shows promise for improving accuracy in high-resolution picture analysis, particularly for applications in disaster relief and urban planning.

V. APPROACHES

1) *R-CNN:* Region-based Convolutional networks do an initial area search before categorizing data, as suggested by their name. Region search is a technique for finding an object in an image. The exhaustive search approach was developed by J.R.R. Uijlings et al. in 2012 as an alternative to selective search, which is one of the strategies for region search. Prior to organizing them into a hierarchy, it first computes the input image's smaller segments or components. [18] The entire image is thus contained in the final group or hierarchical structure. Colour space and similarity metrics are taken into account when sorting the regions that have been found. A predetermined number of region recommendations are generated as a result, and they are then combined to produce the final image. In R-CNN, a selective search approach is utilized to locate area proposals, and deep learning is then applied to locate items in those located region proposals. Then, CNN is used to extend each individual region proposal to fit the CNN's input size, which helps to extract the 4096-dimensional feature vectors created by U. Mittal et al. [12] The probability for each class is predicted using several classifiers that input these feature vectors.

For each class that has pretrained support vector machines (SVM), the likelihood of locating the objects using these feature vectors is forecasted. Linear regression can be used in the region proposal, which alters the bounding box shapes and sizes, to lessen the error in object localization.

2) *Speedy R-CNN:* The Fast Region-based Convolutional Network (Fast R-CNN) was developed in 2015 by R. Girshick et al. [6]. It shares some similarities with RCNN, but its main objective is to shorten the time required to assess each area proposal that is connected to numerous models. Fast R-CNN uses several Convolutional layers and the full image as an input, in contrast to R-CNN, which requires CNN for each region proposal. The area of interest (RoI) is located using a selective search method on CNN's feature maps. With length and width as the main considerations, the RoI pooling layer should be utilized to condense the size of the feature maps in order to produce an accurate region of interest. [15] The output from each individual RoI layer is fed into the fully-connected layers, which then produce a feature vector. These feature vectors and a SoftMax classifier are then used to identify the items, and a linear regression is used to alter their placement.
3) *Faster R-CNN:* S. Ren et al. created the region proposal network (RPN) in 2016 as a new technique for the detection of objects, prediction of bounding boxes, and generation of region proposals in order to address the cost issue in the conventional method, which uses selective search methods for the generation of region proposals. The quicker region proposal follows as a result. Combinations of the region proposal network and rapid region proposal convolutional neural network models are used to compute convolutional neural networks. In order to construct feature maps in this, an entire image is considered as an input. By swiping a window of 3x3 pixels across a whole feature map and linking it to two completely connected layers, a feature vector is created. One layer is used for box classification, and the other layer is used for box regression. Fully connected layers are used to find a lot of region proposals. The output size for the box-regression layer is 4k and the output size for the box-classification layer is 2k when the top limit of k regions is permanently set. They are referred to as anchors when k region suggestions are discovered utilizing a sliding window. Once the anchor boxes were located, only the boxes that were useful after applying a threshold to the objectness were selected. [12] The main CNN model generates feature maps and anchor boxes, which the fast R-CNN model then uses as inputs. Faster R-CNN replaces the selective search strategy with RPN in order to speed up training and testing and enhance performance. For classification, RPN is used to the pre-trained ImageNet dataset, and it is fine-tuned using the PASCAL VOC dataset. In order to quickly train R-CNN, a regional proposal is developed and combined with anchor boxes. It is therefore a time-consuming process.
4) *YOLO:* Bounding box identification and probability prediction are carried out utilizing a single evaluation and a single neural network model in the YOLO technique. Because it delivers real-time forecasts, it is simple to use. The YOLO technique predicts B bounding boxes and their confidence scores using the individual network cells after fragmenting an input image into a SxS network utilizing a whole image as an input. The confidence score can be calculated by combining the likelihood of finding items, the value of IoU within the ground truth boxes, and the predicted. The CNN GoogLeNet model is used in YOLO to display the inception module [9]. There are two fully connected layers, 24 convolutional layers, and either a 1x1, 2x2, or 3x3 filter after that. A 3x3 convolutional layer is used in place of the inception module, and a dimensionality reduction layer comes next. YOLO comes in three different iterations (v1, v2, and v3), with v3 being the most recent and fastest. It includes several filters and nine convolutional layers.

VI. APPLICATIONS

1) *Autonomous cars detection:* In order for autonomous cars to communicate with and understand their surroundings, object detection is crucial. Especially in the case of self-driving autos. Deep learning models can identify pedestrians, vehicles, traffic signs, and other objects on the road in order to make informed decisions, navigate safely, and avoid collisions. The real-time processing of object detection algorithms is essential for the dependability and safety of autonomous driving systems.[12]
2) *Healthcare and Medical Imaging:* In the healthcare sector, deep learning-based object detection is used to evaluate medical images such as X-rays, MRIs, and CT scans. By identifying abnormalities, cancers, or anatomical traits, these models aid in the early detection of diseases and provide physicians with vital details for accurate diagnosis and treatment planning. [12] Object detection is particularly useful in medical imaging for detecting minute, subtle anomalies that would be difficult for the human eye to notice.
3) *Industrial Automation and Quality Control:* Object detection is commonly used in industrial settings to carry out processes for automation and quality control. Manufacturing businesses use deep learning models to inspect products on production lines and find flaws in order to guarantee that only high-quality products are released onto the market. Finding product faults, blemishes, or abnormalities can be a difficult task, but object detecting systems can handle it with ease, considerably boosting manufacturing efficiency and reducing waste.

4) *Virtual reality (VR) and augmented reality (AR) apps:* Object recognition is a crucial aspect of these two categories of apps. By detecting real-world objects and their locations, virtual things can be precisely placed onto the user's environment in augmented reality (AR). [13] The object detection feature in virtual reality (VR) enables users to interact with the virtual environment by identifying and responding to real-world items or gestures. This enhances the immersive experience and opens up new possibilities for games, interactive digital content, and training simulations.
5) *Environmental Monitoring and Conservation:* Deep learning-based object detection has been used for animal preservation and environmental monitoring. Scientists and conservationists use this technology to monitor animal populations, trace down illegal activities like deforestation and poaching, and identify threatened and endangered species. The ability to automatically identify and categorize species using camera trap photos or aerial surveys enables better conservation strategies and biodiversity protection.[13]
6) *Retail analytics and customer information:* Retailers utilize object detection to gather pertinent customer data and enhance store layouts. By analysing customer behaviour, foot traffic, and product interactions, retailers can enhance the shopping experience, optimize product placements, and make data-driven decisions to increase sales and customer happiness. [14] Customers can make purchases without using typical cashiers or self-checkout kiosks by using an object detection-based cashier less checkout system.
7) *Gesture Recognition and Human-Robot Interaction:* Object detection is a key component of gesture recognition systems, which enable both human-computer and human-robot interaction. These technologies enable users to connect easily and naturally with machines, robots, and virtual interfaces by deciphering hand signals and movements. [15] This technology has applications in a variety of fields, including gaming and entertainment as well as assisting those with physical disabilities in using assistive technologies.

VII. FUTURE SCOPE

The future of deep learning-based object detection is promising in many industries. With more research and development, deep learning models like CNNs, RNNs, and transformer-based architectures will become substantially more accurate, quick, and resilient. Greater model generalization will be enabled by larger and more diverse datasets, which will enhance item detection under challenging real-world conditions.

As deep learning infrastructure advances, specialized accelerators like GPUs and TPUs will be paired with edge computing hardware to enable real-time and low-latency object detection. Applications like autonomous vehicles, robots, and surveillance systems will benefit from this. We will have a more complete image of the environment thanks to multi-modal object detection, which is crucial for initiatives like driverless vehicles and smart cities.[17] It accomplishes this by combining data from a variety of sensors, including radar, LiDAR, and cameras. Cross-domain object detection enables models trained in one domain to be applied to other domains with little alteration, reducing the need for labelled data and speeding up the development of object identification systems in new contexts. Combining object identification with other AI specialties like knowledge graph mining and natural language processing (NLP) enables the development of contextually aware systems with enhanced comprehension and more precise detections. Future predictions indicate that deep learning-based object recognition will find widespread application in the environmental monitoring, retail, agricultural, and healthcare sectors. [20] Accurate pest detection in agriculture, improved inventory control in retail, and early disease diagnosis in medical imaging are just a few examples of applications. These advancements will have a big impact on society and technology, making way for safer, smarter systems that can deal with real-world issues.

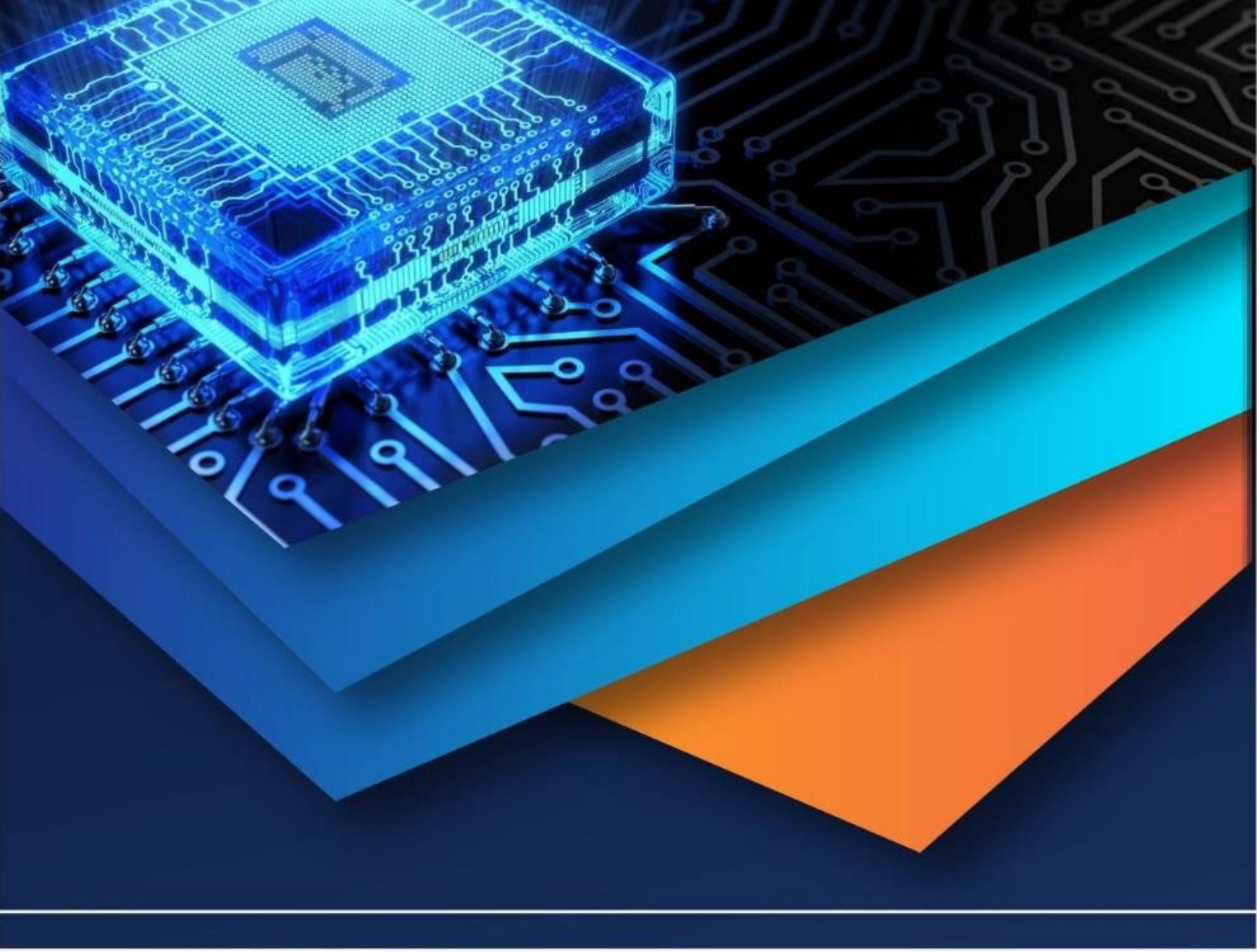
VIII. CONCLUSION

Object identification has been totally altered by deep learning, enabling accurate and efficient object recognition in both still and moving photos. Various techniques, such as R-CNN, Fast R-CNN, Faster R-CNN, and YOLO, have demonstrated appreciable advancements in this field. Despite current progress, there are still unresolved research issues in fields such handling occlusions and small object identification, leaving potential for additional study.

The future of deep learning object recognition is promising thanks to advancements in model architecture and hardware accelerators. These developments will make it feasible to create safer, more efficient, and intelligent systems in industries including healthcare, retail, transportation, and environmental monitoring. We shall be closer to a technologically advanced and interconnected future if we continue to work across disciplines to push the boundaries of object detection.

REFERENCES

[1] G. Chandan, A. Jain, H. Jain, and Mohana, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Jul. 2018
[2] Z.-Q. Zhao, P. Zheng, S.-T. 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, Nov. 2019
[3] S. Goyal and P. Benjamin, "Object Recognition Using Deep Neural Networks: A Survey," arXiv.org, Dec. 10, 2014
[4] A. Boukerche and Z. Hou, "Object Detection Using Deep Learning Methods in Traffic Scenarios," ACM Computing Surveys, vol. 54, no. 2, pp. 1–35, Mar. 2021
[5] W. Budiharto, A. A. S. Gunawan, J. S. Suroso, A. Chowanda, A. Patrik, and G. Utama, "Fast Object Detection for Quadcopter Drone Using Deep Learning," 2018 3rd International Conference on Computer and Communication Systems (ICCCS), Apr. 2018
[6] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable Object Detection Using Deep Neural Networks," 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 2014
[7] Y. P. Chen, Y. Li, and G. Wang, "An Enhanced Region Proposal Network for object detection using deep learning method," PLOS ONE, vol. 13, no. 9, p. e0203897, Sep. 2018
[8] R. Goel, A. Sharma, and R. Kapoor, "Object Recognition Using Deep Learning," Journal of Computational and Theoretical Nanoscience, vol. 16, no. 9, pp. 4044–4052, Sep. 2019
[9] Nita S. patil, Sanjay M. Patil, Chandrashekhar M. Raut, Amol P. Pande, Ajay Reddy Yeruva, and Harish Morwani, "An Efficient Approach for Object Detection using Deep Learning," Journal of Pharmaceutical Negative Results, pp. 563–572, Nov. 2022
[10] J. Kim and V. Pavlovic, "A Shape-Based Approach for Salient Object Detection Using Deep Learning," Computer Vision – ECCV 2016, pp. 455–470, 2016
[11] H. Kim, Y. Lee, B. Yim, E. Park, and H. Kim, "On-road object detection using deep neural network," 2016 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Oct. 2016
[12] U. Mittal, S. Srivastava, and P. Chawla, "Review of different techniques for object detection using deep learning," Proceedings of the Third International Conference on Advanced Informatics for Computing Research, Jun. 2019
[13] M. Shah and R. Kapdi, "Object detection using deep neural networks," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Jun. 2017
[14] U. Subbiah, D. K. Kumar, S. K. Thangavel, and L. Parameswaran, "An Extensive Study and Comparison of the Various Approaches to Object Detection using Deep Learning," 2020 Third International Conference on Smart Systems and Inventive Technology, Aug. 2020
[15] A. Uçar, Y. Demir, and C. Güzeliş, "Object recognition and detection with deep learning for autonomous driving applications," SIMULATION, vol. 93, no. 9, pp. 759–769, Jun. 2017
[16] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," Neurocomputing, vol. 396, pp. 39–64, Jul. 2020
[17] Akshay Mangawati, Mohana, Mohammed Leesan, H. V. Ravish Aradhya, "Object Tracking Algorithms for video surveillance applications" International conference on communication and signal processing (ICCSP), India, 2018, pp. 0676-0680
[18] Manjunath Jogin, Mohana, "Feature extraction using Convolution Neural Networks (CNN) and Deep Learning" 2018 IEEE International Conference On Recent Trends In Electronics Information Communication Technology,(RTEICT) 2018, India
[19] K. L. Masita, A. N. Hasan and S. Paul, "Pedestrian Detection Using R-CNN Object Detector," in 2018 IEEE Latin American Conference on Computational Intelligence (LACCI), Gudalajara, Mexico, Mexico, 2019
[20] Du, J. (2018). Understanding of Object Detection Based on CNN Family and YOLO. Journal of Physics Conference Series 1004(1):012029 , 1-8



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)