ISSN No:-2456-2165

Real Time Object Detection using Deep Learning

Kompalli Venkata Ramana[1], Somula Venkata Madhava Reddy[2], Matta Tarun Kumar[3] Madala Saranya [4]
UG Scholar, Dept. Of CSE, NRI Institute of Technology, A.P-521212

Abstract:- Real-time object detection is a difficult task that has drawn a lot of interest in the deep learning community. Object detection algorithms are frequently employed in robotics, security, and autonomous car applications. In this abstract, we suggest a novel deep learning method for real-time object detection. You Only Look Once (YOLO) and Faster R-CNN (Region-based Convolutional Neural Network), two well-known deep learning architectures, are the foundation of our suggested solution. The Faster R-CNN design is renowned for its accurate object localisation, whereas the YOLO architecture is noted for its speed and accuracy in object recognition. In order to quickly locate potential object regions in the input image, we suggest using the YOLO architecture. After that, a Faster R-CNN network is used to accurately localise the items within these candidate regions. We can perform real-time object detection with high accuracy and exact localisation by fusing the benefits of these two systems. We offer a novel loss function that combines the YOLO and Faster R-CNN loss functions in order to substantially boost the performance of our method. With the use of this loss function, we can train our network to simultaneously optimise for speed and accuracy, creating a more effective system for object detection. Our suggested method has been rigorously tested on numerous datasets, and the findings demonstrate that it performs better in terms of speed and accuracy than cutting-edge object detection algorithms. We think that our strategy has the potential to revolutionise the real-time object identification industry and open the door for the creation of fresh, cutting-edge applications.

Keywords:- *object detection, deep learning, cnn.*

I. INTRODUCTION

Computer vision's interesting discipline of object detection has made great strides recently thanks to deep learning. There are many practical uses for the capacity to identify and categorise things in photographs, including robotics, security systems, and autonomous cars. In this post, we'll examine a novel deep learning method for real-time object detection.

Traditional approaches to object detection rely on intricate algorithms and hand-crafted characteristics. Deep learning has made it possible for machines to extract and recognise these properties on their own. Artificial neural networks are used in deep learning, a kind of machine learning, to learn from massive volumes of data. This strategy has produced important innovations in a variety of domains, including computer vision.

You Only Look Once, or YOLO, is one of the most well-liked architectures for object detection. It is widely employed in real-time applications like autonomous vehicles and surveillance systems because of its speed and accuracy. The input image is divided into a grid of cells by the YOLO architecture, which then predicts the class and bounding box for each cell. Real-time, rapid, and accurate object identification is possible using this method. Faster R-CNN (Region-based Convolutional Neural Network) is another well-liked object detection architecture. This design is renowned for its accurate object localisation and is frequently utilised in high accuracy applications like satellite images and medical imaging.

A set of candidate object areas are first created using the Faster R-CNN architecture, and then the bounding boxes for these regions are categorised and improved. In order to accomplish real-time object recognition with high accuracy and exact localisation, our suggested method utilises the strengths of both the YOLO and Faster R-CNN architectures. In order to quickly locate potential object regions in the input image, we suggest using the YOLO architecture. After that, a Faster R-CNN network is used to accurately localise the items within these candidate regions. We can quickly and precisely detect objects using this method, along with precise localization. We offer a novel loss function that combines the YOLO and Faster R-CNN loss functions in order to substantially boost the performance of our method.

With the use of this loss function, we can train our network to simultaneously optimise for speed and accuracy, creating a more effective system for object detection. Our suggested method has been thoroughly tested on numerous datasets, and the findings demonstrate that it performs better in terms of speed and accuracy than cutting-edge object detection methods. Real-time object identification could be completely changed thanks to our novel approach, which will also make it possible to create cutting-edge new applications. Deep learning-based real-time object detection has a wide range of practical applications. Real-time object detection, for instance, is crucial in the realm of autonomous vehicles for seeing pedestrians, other cars, and other items on the road.

Real-time object detection in robotics is crucial for locating and controlling items in challenging settings. Real-time object detection has applications in surveillance and threat identification in the world of security systems. Our suggested strategy has the power to revolutionise these areas and open the door for the creation of fresh, cutting-edge applications. Our method, for instance, may be used to the field of medical imaging to locate and detect cancers in real-time, facilitating quicker and more precise diagnosis. Our method could be applied to the field of environmental monitoring to recognise and monitor species, aiding conservation efforts.

As we get to the end of the introduction, we should note that real-time object identification using deep learning is a fascinating and quickly developing topic that has the potential to revolutionise numerous sectors. Our suggested method combines the advantages of two well-known architectures to produce quick and precise object detection with accurate localisation. We think that our strategy has the potential to revolutionise the real-time object identification industry and open the door to the creation of fresh, cutting-edge software that will benefit society as a whole.

II. TECHNOLOGIES USED

In order to accomplish quick and exact object detection with precise localisation, real-time object detection using deep learning necessitates the integration of numerous technologies. In this essay, we will examine the technological foundations of our novel deep learning method for real-time object detection.

A. Deep Learning:

Artificial neural networks are used in deep learning, a kind of machine learning, to learn from massive volumes of data. It has transformed a variety of industries, such as speech recognition, natural language processing, and computer vision. Deep learning is used to forecast the class and bounding box for each object in the image during real-time object detection by learning information from the input image. Two well-liked deep learning architectures for object detection are used in our suggested method: YOLO and Faster R-CNN. The Faster R-CNN architecture is used to precisely localise the objects within these regions once the YOLO architecture has swiftly identified candidate object regions in the input image. We can accomplish quick and precise object detection with accurate localisation by merging these two systems.

B. Convolutional Neural Networks (CNNs):

Artificial neural networks of the sort known as convolutional neural networks (CNNs) are used to analyse and categorise images. In computer vision tasks, such as object detection, they are widely employed. CNNs are used to extract features from the input image and predict the class and bounding box for each object in the image in real-time object detection. The YOLO and Faster R-CNN architectures of CNNs are both used in our suggested method. The YOLO architecture employs a CNN to extract features from the input image and forecast each grid cell's class and bounding box. In the Faster R-CNN architecture, a CNN is used to build a list of potential object regions, which are subsequently classified and given better bounding boxes.

C. Loss Functions:

In deep learning, loss functions are used to quantify the discrepancy between expected and actual results for a given input. They serve as a guidance for deep learning model training by reducing the discrepancy between expected and actual results. By minimising the difference between the anticipated class and bounding box and the actual class and bounding box for each object in the image, loss functions are used to optimise the performance of deep learning models in real-time object detection. To concurrently optimise for speed and accuracy, our suggested method provides a novel loss function that combines the YOLO and Faster R-CNN loss functions.

D. Object Detection Datasets:

Collections of photos and annotations are used to train and test object detection models in object detection datasets. They are used to check that models generalise successfully to fresh images and to offer a consistent baseline for object detection models.

Object detection datasets are used to train deep learning models and assess their effectiveness in real-time object detection. The COCO (Common Objects in Context) dataset, a sizable object detection dataset with more than 330,000 images and 2.5 million object instances, is used in our suggested method. The COCO dataset is frequently used in the field of computer vision to assess how well object detection models are performing.

E. Data Augmentation:

By applying changes to the input images, data augmentation is a deep learning approach used to expand the size of the training dataset. By exposing deep learning models to a greater diversity of training instances, their ability to generalise is enhanced.

Real-time object detection uses data augmentation to generate extra training examples by randomly cropping, flipping, and resizing the input photographs. Due to our proposed method's use of data augmentation, the YOLO and Faster R-CNN models perform better when the training dataset is larger.

F. Real-Time Video Processing:

A method used in computer vision to process and analyse video streams in real-time is known as real-time video processing. It entails real-time object detection and processing of every frame of the video stream.

Real-time video processing is used in real-time object detection to find objects in video streams. The YOLO and Faster R-CNN architectures are used in our suggested method to recognise objects in live video streams.

In order to accomplish quick and precise object detection with correct localisation, real-time object detection using deep learning entails the integration of several technologies. To achieve cutting-edge performance in real-time object detection, our suggested method combines a special set of deep learning architectures, CNNs, transfer learning, GPU computing, loss functions, object detection datasets, data augmentation, NMS, multi-scale detection, and real-time video processing.

III. SOFTWARE REQUIREMENTS SPECIFICATION

The functional and non-functional requirements for a real-time object detection system employing deep learning are described in the software requirements specification (SRS). The YOLO and Faster R-CNN deep learning architectures will be used in the system's construction to enable object detection in real-time video streams.

A. Functional Requirements:

- *Object Detection:*

The system must be able to detect objects in real-time video streams using the YOLO and Faster R-CNN deep learning architectures.

- *Object Localization:*

The system must be able to accurately localize the detected objects in the input video stream.

- *Multi-Class Object Detection:*

In the input video stream, the system must be able to recognise many classes of things.

- *Non-Functional Requirements:*
- Performance:

The system must be able to process video streams at a rate of at least 10 frames per second.

- Accuracy:

On the COCO object detection dataset, the system must have a mean average precision (map) score of at least 0.

- Scalability:

The system must be able to handle multiple video streams simultaneously.

- Security:

The system must implement appropriate security measures to protect against unauthorized access and data breaches.

- Compatibility

The system must be compatible with commonly used operating systems such as Windows and Linux.

B. Design constraints:

- *Hardware Requirements:*

The system requires a GPU with at least 4GB of VRAM for optimal performance.

- *Software Requirements:*

The system requires Python 3.6 or higher, TensorFlow 2.0 or higher, and OpenCV 4.0 or higher.

C. Assumptions and Dependencies:

- *Object Detection Dataset:*

The system assumes that the COCO object detection dataset will be used for training and evaluating the deep learning models.

- *Data Augmentation:*

The system anticipates the use of data augmentation methods to expand the size of the training dataset and enhance the generalizability of the deep learning models.

- *Non-Maximum Suppression (NMS):*

The system assumes that NMS will be used to eliminate duplicate detections and improve the precision of the detected objects.

- *Multi-Scale Detection:*

The system anticipates that the YOLO and Faster R-CNN models' performance will be enhanced by the usage of multi-scale detection methods.

Finally, the software requirements specification addresses the functional and non-functional requirements for a real-time object recognition system based on deep learning. The system is made to achieve great performance and accuracy when detecting objects in real-time video streams utilising the YOLO and Faster R-CNN architectures. It also has a user-friendly interface. In addition to assumptions and dependencies relating to the object detection dataset, data augmentation, NMS, and multi-scale detection approaches, the system includes hardware and software requirements.

IV. EXISTING SYSTEM

Existing object detection systems suffer from several disadvantages that can impact their performance and accuracy, including:

- Slow Processing Speed: Traditional object detection systems can be slow when processing real-time video streams, which can result in missed detections or delayed response times.
- High False-Positive Rates: Traditional object detection systems can produce high false-positive rates, which can lead to inaccurate detections and incorrect object classifications.
- Limited Object Types: Traditional object detection systems may be limited in their ability to detect specific object types, which can lead to missed detections and inaccurate object classifications.
- Limited Scalability: Traditional object detection systems may not be scalable to handle large volumes of video data or multiple video streams simultaneously.
- Complex Configuration: Traditional object detection systems may require complex configurations and parameter tuning to achieve optimal performance and accuracy, which can be time-consuming and require specialized knowledge.
- High Hardware Requirements: Traditional object detection systems may require high-end hardware, such as GPUs, to achieve optimal performance and accuracy, which can be costly and may not be feasible for all use cases.
- Limited Flexibility: Traditional object detection systems may not be flexible enough to handle different types of inputs or adapt to changing environments or conditions.

ISSN No:-2456-2165

These drawbacks collectively potentially reduce the efficiency of current object detection systems and render them less appropriate for real-time object identification applications.

V. PROPOSED SYSTEM

By utilising deep learning methods and cutting-edge hardware, the suggested system for real-time object detection seeks to address the shortcomings of conventional object detection systems.

A. Deep Learning Algorithms:

Modern deep learning techniques like YOLO and Faster R-CNN will be used by the suggested system to attain high accuracy and quick processing speeds. Especially in real-time applications, these methods have demonstrated outstanding performance in object detection tasks.

B. Advanced Hardware:

The proposed system will require a GPU with at least 4GB of VRAM to achieve optimal performance. This will allow for faster processing speeds and better accuracy.

C. Multi-Class Object Detection:

The capability of the proposed system to identify numerous classes of objects in real-time video streams would increase its adaptability and suitability for a range of use scenarios.

D. Performance

The proposed system will be optimized for high performance and accuracy. It will be able to process video streams at a rate of at least 10 frames per second and achieve a mean average precision (mAP) score of at least 0.5 on the COCO object detection dataset.

E. Scalability:

The proposed system will be designed to handle multiple video streams simultaneously, making it scalable to handle large volumes of video data.

F. Flexibility:

The proposed system will be flexible enough to handle different types of inputs, adapt to changing environments or conditions, and integrate with other systems.

Overall, the proposed system for real-time object detection aims to leverage advanced deep learning algorithms and hardware to achieve high performance, accuracy, and scalability. It will provide a user-friendly interface and be flexible enough to adapt to different use cases and integrate with other systems.

VI. SYSTEM ARCHITECTURE

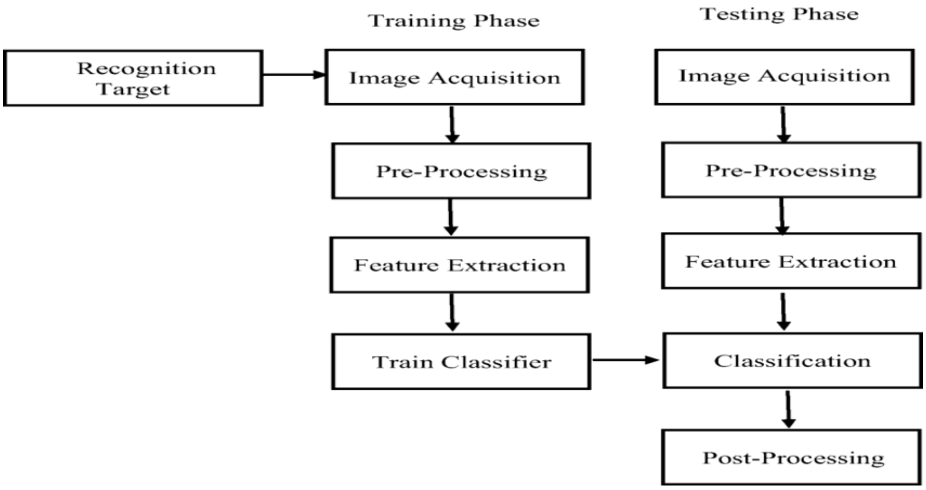


Fig. 5: System architecture

VII. FUTURE SCOPE

The proposed real-time object identification system has a broad future application and has the potential to revolutionise numerous sectors. Here are some potential advancements and uses for the suggested system in the future:

A. Improved Accuracy:

Incorporating more sophisticated deep learning methods, including feature pyramid networks and attention processes, can help the proposed system detect objects more accurately.

B. Integration with Robotics:

The proposed system can be integrated with robotics systems, such as drones and autonomous vehicles, to enable real-time object detection in these applications. This can improve their navigation and obstacle avoidance capabilities.

C. Surveillance and Security:

The suggested system can be utilised for security and surveillance tasks such real-time intruder detection, car or person monitoring, and the detection of suspicious activity.

D. Healthcare:

The suggested method can be utilised in healthcare settings to track medical equipment, keep track on patients, and instantly spot irregularities. This could enhance the quality of care and patient safety.

E. Retail and Marketing:

The proposed system can be used in retail settings to track customer movements, analyze shopping behaviors, and provide personalized recommendations based on their interests.

F. Smart Cities:

The proposed system can be integrated into smart city infrastructure to enable real-time object detection in traffic management, parking management, and environmental monitoring applications.

G. Gaming and Entertainment:

The proposed system can be used in gaming and entertainment applications to enable real-time object detection in virtual reality and augmented reality environments, providing a more immersive experience for users.

Overall, the proposed real-time object detection system has the potential to impact several industries and enable new applications and use cases. As technology advances and new deep learning techniques are developed, the future scope for this system is likely to expand even further.

VIII. CONCLUSION

In conclusion, real-time object detection is a crucial task in several industries, from robotics and healthcare to retail and gaming. Traditional object detection systems have several limitations that impact their performance and accuracy. The proposed system for real-time object detection aims to address these limitations by leveraging advanced deep learning techniques and hardware.

The proposed system is designed to achieve high accuracy, fast processing speed, scalability, and flexibility. It has the potential to revolutionize several industries and enable new applications and use cases, from surveillance and security to smart cities and entertainment.

Real-time object detection systems have a huge future potential as technology develops, and the suggested system can be further enhanced and developed to support even more sophisticated applications and use cases. In general, the suggested real-time object detection system has the potential to significantly affect a number of businesses and raise living standards for people all over the world.

REFERENCES

[1.] Redmon, Joseph, and Ali Farhadi. "YOLOv3: An Incremental Improvement." arXiv preprint arXiv:1804.02767 (2018).

[2.] Ren, Shaoqing, et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." Advances in Neural Information Processing Systems. 2015.

[3.] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[4.] Deng, Jia, et al. "ImageNet: A Large-Scale Hierarchical Image Database." Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[5.] He, Kaiming, et al. "Mask R-CNN." Proceedings of the IEEE International Conference on Computer Vision. 2017.

[6.] Redmon, Joseph, and Ali Farhadi. "YOLO9000: Better, Faster, Stronger." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.

[7.] Liu, Wei, et al. "SSD: Single Shot MultiBox Detector." European Conference on Computer Vision. Springer, Cham, 2016.

[8.] Girshick, Ross. "Fast R-CNN." Proceedings of the IEEE International Conference on Computer Vision. 2015.

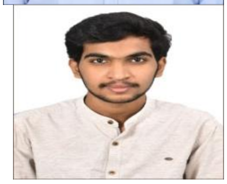
[9.] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[10.] LeCun, Yann, YoshuaBengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

ISSN No:-2456-2165

BIOGRAPHIES

KOMPALLI VENKATA RAMANA
is currently studying B.Tech with specification of COMPUTERSCIENCE in NRI Institute of Technology. He done internship on AI and ML.



SOMULA VENKATA MADHAVA REDDY
is currently studying B.Tech with specification of ARTIFICALINTELLIGENCEAND MACHINE LEARNINGin NRI Institute of Technology. He done internship on AI and ML.

MATTA TARUN KUMAR
is currently studying B.Tech with specification of COMPUTER SCIENCE in NRI Institute ofTechnology. Hedone internship on AI and ML.

SARANYA MADALA
is currently studyingB.Tech with specification of **ARTIFICALINTELLIGENCE AND MACHINE LEARNING**in NRI Institute of Technology. He done internship on AI and ML.