International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Advanced Deep Learning for Real-Time Object Identification and Assistance for the Blind

BUVANESWARI S[1],POOJA A[2],PRATHAP A[3],PREMKUMAR B[4],Mrs.PREMADEVI P[5]

1,2,3,4 B.E, Student Department Computer Science And Engineering, Angel College Of Engineering And Technology, Tiruppur, Tamil Nadu, India.

5 Asst. Professor, Department Computer Science And Engineering, Angel College Of Engineering And Technology, Tiruppur, Tamil Nadu, India.

ABSTRACT :

" Advanced Deep Learning for Real-Time Object Identification and Assistance for the Blind" and is developed using HTML, CSS, and JavaScript for the front end, with Python (Flask framework) for the back end. Object Detection involves identifying and recognizing real-world objects such as backgrounds, airplanes, bicycles, birds, boats, bottles, buses, cars, cats, chairs, cows, dining tables, dogs, horses, motorbikes, people and humans from images or videos. This technique enables understanding the details of a webcam recognizing, localizing, and detecting multiple objects within it. Cater to various real-world applications. Webcam input allows users to process webcam object for object detection, analyzing each frame in real time Camera input integrates live camera feeds, enabling users to detect objects in their surroundings in real time, making it highly useful for tasks requiring immediate feedback, such as assisting visually impaired individuals. This system is motivated by the need to aid visually impaired individuals, those who face challenges in navigating unfamiliar environments. By utilizing real-time object detection through camera inputs, our system aims to assist visually impaired users by identifying objects around them, and also get a voice of the object thus enhancing their ability to navigate independently.

INDEX TERMS: Real-time object detection, Deep learning, Computer vision

I. INTRODUCTION:

Object detection is the process of identifying and recognizing instances of real-world objects such as backgrounds, airplanes, bicycles, birds, boats, bottles, buses, cars, cats, chairs, cows, dining tables, dogs, horses, motorbikes, people, potted plants, sheep, sofas, trains, monitors, and humans from images or videos. It enables the understanding of an image or video by allowing recognition, localization, and detection of multiple objects within a scene. This technology is a foundational aspect of computer vision, which plays a crucial role in various applications ranging from autonomous vehicles to healthcare.

The significance of real-time object detection cannot be overstated, especially in applications requiring immediate responses, such as autonomous driving, security surveillance, industrial automation, and assistive technologies for the visually impaired. The proposed system, designed specifically to assist visually impaired individuals, leverages advanced deep learning algorithms to deliver accurate and efficient object identification in real-time.

Unlike traditional methods, which rely on handcrafted features, this system utilizes deep neural networks to automatically learn and identify object features from images or video streams. This approach ensures higher accuracy, even in complex and dynamic environments. The system is designed to be highly adaptable, capable of handling various input types, including static images, video streams, and live camera feeds.

One of the primary objectives of this project is to ensure a user-friendly interface, making the system accessible to users with limited technical expertise. The system's flexibility allows it to be easily integrated with other applications or platforms, expanding its range of use cases beyond assisting the visually impaired. It can be deployed in smart cities, retail, manufacturing, and other sectors where real-time object detection is valuable.

As we advance further into the era of artificial intelligence, real-time object identification using deep learning stands out as a rapidly evolving field. Our proposed approach combines the strengths of two well-known architectures to achieve fast and accurate object detection with precise localization. We believe that this system has the potential to transform real-time object detection, offering a robust and scalable solution for diverse applications.")

II.LITERATURE SURVEY:

TRANSFER LEARNING FOR IMPROVED OBJECT DETECTION PERFORMANCE:

Azizpour et al.

Transfer learning utilizes pre-trained deep learning models, originally trained on large-scale image datasets like ImageNet, to accelerate and enhance the training of object detectors on new, often smaller, datasets. This method mitigates the need for extensive labeled data in specialized domains, improving generalization and reducing overfitting. Recent advances combine transfer learning with fine-tuning strategies that selectively retrain deeper network layers to adapt to domain-specific features, thereby achieving better detection accuracy and faster convergence. Such approaches have proven effective in medical imaging, autonomous driving, and assistive technology applications, where annotated data can be scarce or expensive to obtain.

ATTENTION MECHANISMS TO ENHANCE FEATURE REPRESENTATION IN DETECTION NETWORKS:

Hu, Shen, and Sun

Attention modules, particularly the Squeeze-and-Excitation (SE) block, dynamically adjust the importance of each feature channel by learning adaptive weights during training. This channel-wise attention improves the network's ability to focus on salient object features while suppressing irrelevant background noise. Beyond SE blocks, other attention mechanisms such as spatial attention and self-attention (used in Transformer-based models) have demonstrated further improvements in object localization and classification. Incorporating attention not only enhances accuracy but also enables models to be more resilient to occlusions, cluttered backgrounds, and varying lighting conditions, which are common challenges in real-world detection scenarios.

MULTI-SCALE FEATURE EXTRACTION FOR ROBUST OBJECT DETECTION:

Liu et al.

Effective detection of objects at multiple scales remains a significant challenge due to varying object sizes and aspect ratios. Feature Pyramid Networks (FPN) address this by building a top-down architecture with lateral connections to combine high-resolution, low-level features with low-resolution, high-level semantic features. This fusion enriches the feature maps used for detection, improving small object detection without sacrificing speed. Multi-scale training and data augmentation strategies further complement FPNs, helping models learn scale-invariant representations. These techniques are vital for applications such as traffic monitoring and assistive vision devices, where objects of interest may appear at different distances and sizes.

OBJECT DETECTION IN VIDEO STREAMS WITH TEMPORAL CONSISTENCY:

Song et al.

Video object detection extends beyond individual frame analysis by incorporating temporal context to improve accuracy and reduce flickering detections. Methods combine convolutional neural networks (CNNs) with temporal modules such as Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, or optical flow to capture motion and maintain object identity over time. These models improve robustness against challenges such as motion blur, sudden illumination changes, and occlusions. Additionally, real-time video detection frameworks leverage temporal smoothing and tracking-by-detection approaches, crucial for autonomous vehicles, robotics, and security surveillance systems.

EDGE COMPUTING AND LIGHTWEIGHT MODELS FOR REAL-TIME OBJECT DETECTION:

Li et al.

Edge computing brings computation closer to data sources, reducing latency and bandwidth constraints inherent in cloud-based systems. Lightweight deep learning architectures such as MobileNet, ShuffleNet, and EfficientDet have been optimized to run efficiently on edge devices including smartphones, embedded systems, and IoT devices. These models employ depthwise separable convolutions, neural architecture search, and compound scaling techniques to balance speed and accuracy. The deployment of such models in assistive technology enables on-device real-time object detection with privacy preservation, minimal power consumption, and instant response—critical factors for applications supporting visually impaired users or autonomous navigation in constrained environments.

III. RELATED WORK :

Object detection has been extensively studied, progressing from traditional handcrafted feature approaches to advanced deep learning techniques. Early works such as Dalal and Triggs [1] introduced Histogram of Oriented Gradients (HOG) features combined with Support Vector Machines (SVM) for object recognition, laying foundational work for machine vision. However, these approaches struggled with complex backgrounds and real-time constraints.

The advent of deep learning revolutionized object detection, beginning with the Region-Based Convolutional Neural Network (R-CNN) introduced by Girshick et al. [2]. Although R-CNN improved accuracy by applying CNNs to selective region proposals, it was computationally expensive. Subsequent models, Fast R-CNN [3] and Faster R-CNN [4], optimized the pipeline by integrating Region Proposal Networks, enhancing both speed and accuracy. Despite these improvements, these two-stage detectors remain resource-intensive for real-time applications.

Single-stage detectors like YOLO [5] and SSD [6] addressed this by directly predicting bounding boxes and class probabilities in one pass, significantly improving inference speed. YOLO's grid-based approach enabled real-time detection, with newer versions such as YOLOv4 [7] and YOLOv7 [8] achieving state-of-the-art accuracy and efficiency. SSD's multi-scale feature maps further improved detection of objects at different sizes, making it suitable for various real-world scenarios.

Lightweight architectures such as MobileNet [9] and EfficientDet [10] have been developed to run efficiently on mobile and embedded devices. These models are particularly important for applications in assistive technology where computational resources and power consumption are limited. Edge computing solutions leverage these lightweight models to perform real-time detection directly on devices, reducing latency and improving privacy [13].

Attention mechanisms have been incorporated into detection networks to improve feature representation, allowing models to better focus on relevant spatial and channel-wise information [14].

IV. EXISTING SYSTEM:

Existing object detection systems suffer from several disadvantages that can impact their performance and accuracy, 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 of the 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. The systems may not be flexible enough to handle different types of inputs or adapt to changing environments or conditions.

Disadvantages of the Existing System:

- Existing object detection systems can be slow when processing real-time video streams, which can result in missed detections or delayed response times.
- Traditional object detection systems can produce high false-positive rates, which can lead to inaccurate detections and incorrect object classifications.

- The systems may be limited in their ability to detect specific object types, which can lead to missed detections and inaccurate object classifications.
- The object detection systems may not be scalable to handle large volumes of video data or multiple video streams simultaneously.

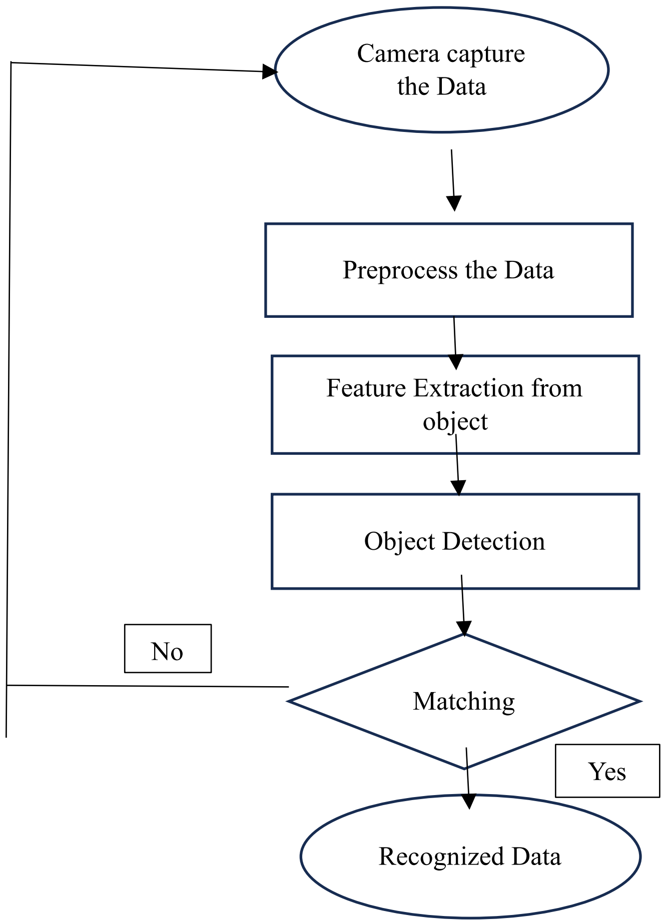
V.PROPOSED SYSTEM:

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

Advantages of the Proposed System:

- Modern deep learning techniques like YOLO will be used by the suggested system to attain high accuracy and quick processing speeds.
- 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
- The proposed system will be designed to handle multiple video streams simultaneously, making it scalable to handle large volumes of video data.

VI.ARCHITECTURE DIAGRAM:



VII.TECHNIQUES:

The proposed system employs advanced deep learning techniques to achieve real-time object detection and voice feedback for visually impaired users. At its core, the You Only Look Once (YOLO) algorithm is used for rapid and accurate object detection from live webcam input. The system integrates lightweight models such as YOLOv4-tiny and Mobile Net to ensure efficient performance on edge devices like Raspberry Pi or mobile phones. These models use convolutional neural networks (CNNs) to extract meaningful features from images. Text-to-Speech (TTS) technology is implemented to provide immediate voice output, converting detected object labels into spoken words. Multilingual support is also integrated to accommodate users from different language backgrounds. Edge computing ensures low latency and fast response without relying on cloud servers. The system utilizes pretrained models and data augmentation techniques to improve accuracy and generalization. Temporal consistency is maintained to reduce flickering detections during movement. Overall, the combination of these techniques results in an assistive system that is fast, portable, and user-friendly.

VIII. PROJECT DESCRIPTION:

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 generalize successfully to fresh images and to offer a consistent baseline for object detection models.

Data Preprocessing:

By applying changes to the input images, data preprocessing is a deep learning approach used to expand the size of the training dataset. YOLO enjoys the benefit of being a lot quicker than different networks nevertheless keeps up with exactness.

Real-Time Video Processing:

A method used in computer vision to process and analyze 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.

Model Training & Evaluation:

A weights file is the final output after training which will be used for object detection in our model. On several videos, the proposed system is tested. Once training is completed the weights files in used for object detection in videos. It also has a user-friendly interface.

IX.CONCLUSION AND FEATURES:

The 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. R. I. Tiyar and D. H. Fudholi, “KajianPengaruh Dataset dan Bias Dataset terhadap Performa AkurasiDeteksiObjek,”Petir, vol.14, no. 2, pp. 258–268, 2021, doi: 10.33322/petir.v14i2.1350.
2. Nurdin, D.Hamdhana, and M. J. Setiawan, “SistemPendeteksiPolaLafadz Allah Dan Muhammad Pada Citra Al-Qur’an MenggunakanMetodePeirce,”e-Journal TechsiTek. Inf., vol. 9, no. 2, pp. 78–90, 2017.
3. Y. Mujiaty, D.Novayanti, H. Khotimah, and Hanifah, “PENGOLAHAN DATA AKUNTANSI MENGGUNAKAN SISTEM ZAHIR VERSI 5.1 PADA PT.MENARA WARNA INDONESIA JAKARTA,”AKRAB JUARA, vol. 5, no. 2, 2020.
4. G. V. Shalini, M. K. Margret, M. J. S. Niraimathi, and S. Subashree, “Retraction: Social Distancing Analyzer Using Computer Vision and Deep Learning,”Journal of Physics: Conference Series, vol. 1916, no. 1. 2021, doi: 10.1088/1742-6596/1916/1/012039.
5. L. Agustien et al., “Real-time Deteksi Masker Berbasis Deep Learning menggunakanAlgoritma CNN YOLOv3,” vol. 8, no. 2, pp. 129–137, 2021.
6. J. R. Parker, Algorithms for image processing and computer vision. John Wiley & Sons, 2010.
7. F. Leite, A. Akcamete, B. Akinci, G. Atasoy, and S. Kiziltas, “Analysis of modeling effort and impact of different levels of detail in building information models,” Autom. Constr., 2011, doi: 10.1016/j.autcon.2010.11.027.
8. Roihan, P. A. Sunarya, and A. S. Rafika, “Pemanfaatan Machine Learning dalamBerbagaiBidang: Review paper,”IJCIT (Indonesian J. Comput. Inf. Technol., vol. 5, no. 1, pp. 75–82, 2020, doi: 10.31294/ijcit.v5i1.7951.
9. Q. C. Mao, H. M. Sun, Y. B. Liu, and R. S. Jia, “Mini-YOLOv3: Real-Time Object Detector for Embedded Applications,”IEEEAccess, vol. 7, 2019, doi: 10.1109/ACCESS.2019.2941547.
10. Z. Zheng, J. Zhao, and Y. Li, “Research on Detecting Bearing-Cover Defects Based on Improved YOLOv3,”IEEE Access, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3050484.