International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 12 | December - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

Object Recognition in AI

[1]Dr. K. Madan Mohan, [2]M. Lalitha Sriya, [3]M Nikitha, [4]M. Surya Teja, [5]K. Srinivas

[1]Assistant Professor, [2,3,4,5]Students of Dept. of Computer Science & Engineering

Sreyas Institute of Engineering and Technology

ABSTRACT

Object Recognition Using AI is a pioneering concept at the crossroads of artificial intelligence and computer vision, devoted to automating the identification and localization of objects within images and video streams. This paradigm harnesses machine learning algorithms to emulate human visual cognition, empowering systems to discern and precisely pinpoint diverse objects, transcending traditional image analysis. The essence of Object Recognition Using AI lies in its transformative impact across a multitude of applications. In retail, it revolutionizes inventory management by autonomously recognizing products on shelves. Within the security domain, it enhances surveillance systems, enabling real-time tracking of individuals and objects. In the automotive landscape, it empowers self-driving vehicles to interpret their surroundings, safeguarding passengers and pedestrians alike. Our research endeavors to advance the field of object recognition through the implementation of Mask R-CNN, a state-of-the-art deep learning architecture. We aim to achieve highly accurate object detection and segmentation by leveraging the power of convolutional neural networks. In pursuit of this, we conducted comprehensive experiments and evaluations. Our results showcase the effectiveness of Mask R-CNN in various real-world scenarios, with impressive precision and recall rates. In conclusion, this study contributes valuable insights and tools to the realm of computer vision, enabling enhanced object recognition applications.

Keywords: Object Recognition, Mask R-CNN, Deep Learning, Convolutional Neural Networks, Computer Vision.

I. INTRODUCTION

An object can be defined as a distinct and identifiable entity or item within an image or a scene, often possessing specific visual characteristics that distinguish it from its surroundings. In an era marked by an unprecedented surge in digital imagery, the demand for automated and precise object recognition has gained paramount importance. The ability to discern and categorize objects within images and videos is not merely a technological feat; it is a pivotal enabler for a plethora of applications spanning from autonomous vehicles to surveillance systems[1].

In this context, our research embarks on a compelling journey into the realm of object recognition, seeking to harness the transformative potential of Mask R-CNN, a cutting-edge deep learning architecture. This study aspires to address the contemporary challenges in object recognition, enhance its precision, and facilitate its adoption across diverse industries. Object detection in artificial intelligence (AI) represents a

sophisticated and integral aspect of computer vision, a field focused on empowering machines to interpret and understand visual information. The primary goal of object detection is to enable machines to identify and locate multiple objects within digital images or video frames. This goes beyond the realm of simple image classification, which assigns a single label to an entire image, by providing a more granular understanding of the visual content.

Object detection involves not only recognizing the presence of objects but also determining their specific locations and spatial extents within the given visual data. This capability is fundamental for a myriad of applications across diverse industries.

One of the key challenges in object detection lies in the variability and complexity of real-world visual data. Images can contain objects of different sizes, orientations, and occlusions, making it essential for object detection models to exhibit robustness and adaptability. In recent years, the field has witnessed remarkable progress, largely driven by the advent of deep learning techniques. Convolutional Neural Networks (CNNs)[2], in particular, have emerged as powerful tools for object detection, demonstrating unprecedented levels of accuracy.

Deep learning models excel at automatically learning hierarchical features from data, allowing them to discern intricate patterns and representations within images. This ability is crucial for distinguishing objects belonging to various classes, thereby facilitating the development of systems capable of recognizing and categorizing objects with remarkable precision. Object detection, powered by these advanced AI models, has found applications in a wide array of domains. For instance, in autonomous vehicles[3], object detection is vital for identifying pedestrians, other vehicles, and obstacles, contributing to enhanced safety and decision-making. Surveillance systems leverage object detection to monitor and analyze video feeds, identifying suspicious activities or objects in real-time. Additionally, object detection plays a pivotal role in augmented reality experiences, where virtual objects need to be seamlessly integrated into the user's physical environment.

As the capabilities of object detection in AI continue to evolve, the technology is becoming increasingly indispensable for automating tasks, improving efficiency, and bolstering safety across various industries. The ongoing research and development in this field promise even greater strides in accuracy, speed, and adaptability, opening up new possibilities for applications that require a nuanced understanding of the visual world.

II. LITERATURE SURVEY

Alex et al. (2019) [4] introduces the AlexNet architecture, which was pivotal in demonstrating the effectiveness of deep convolutional neural networks (CNNs) for image classification. Object recognition often relies on CNNs like these. Authors: Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik (2021)"R-CNNs for Object Detection". This paper discusses the Region-based Convolutional Neural Networks (R-CNN) approach to object detection. It's aprecursor to Faster R-CNN and provides a foundation for the development of modern object detection models.

Shaoqing et al.(2014)[6-10] "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" This paper introduces the Faster R-CNN architecture, which is a widely adopted model for

object detection. It presents the concept of Region Proposal Networks (RPN) for efficient and accurate object detection.

Joseph et al.(2017)[11-13]YOLO: Real-Time Object Detection": The You Only Look Once (YOLO)framework is known for real-time object detection. This paper describesYOLO's single-shot approach, making it suitable for applications that require low-latency object recognition.

Tsung et al. [14-15] introduces the Common Objects in Context (COCO) dataset, a widely used benchmark for object detection and image segmentation. It's essential for evaluating and benchmarking object recognition models.

Gowroju et al.[16-21] introduces Machine Learning and Deep Neural Networks, an approaching recognition for different use-cases. Its scalability is wide ranging and development for various aspects of this project.

2.1 Computer Vision

Computer vision (CV)[4] is a subfield of computer science that focuses on giving computers the ability to understand visual data. Gerald Jay Sussman was given the task by Martin Minsky in the early 1970s or late 1960s to connect a computer to a camera and have the device report what it saw. The field of study known as "computer vision"[1] (CV) strives to develop methods that will enable computers to "see" and decipher the content of digital images such as pictures and movies. It appears to be simple since everyone, even very young toddlers, can figure out the computer vision problem.

Despite this, the issue is still largely unresolved due to both a lack of knowledge on biological vision and the intricacy of visual perception in a dynamic physical universe that is practically infinitely changing.

2.2 Computer Vision vs AI: Studying computer vision focuses on making it easier for machines to see, as demonstrated in figure. It is a field of study that falls under the broad categories of machine learning and artificial intelligence, and it may make use of both specialized methods and general learning algorithms.

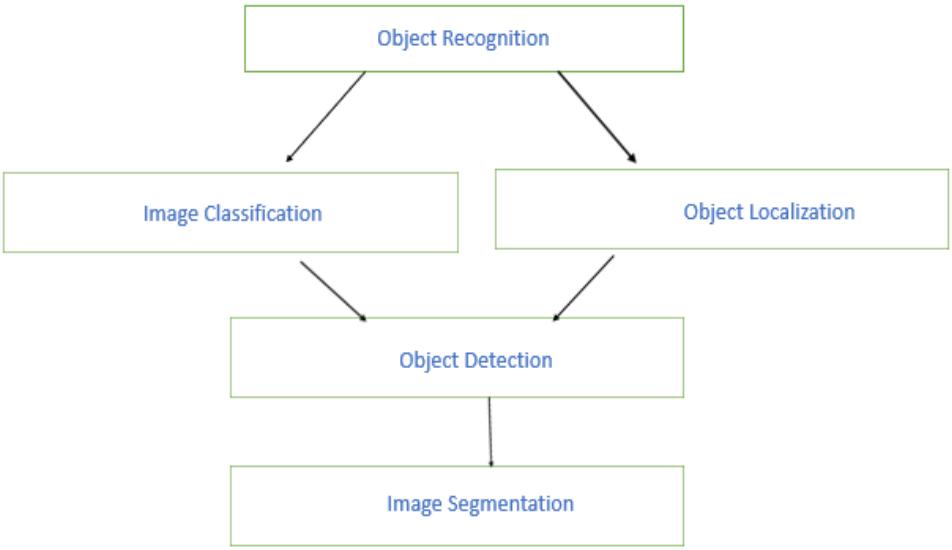

Fig-1:Object Detection Flow

2.3 Computer Vision and Image Processing

Computer vision and image processing are not the same thing. The act of transforming an old image into a new one, usually by improving or streamlining the information is referred to as image processing in some contexts. It's a type of digital signal processing that doesn't care how images are interpreted. For a given computer vision system, image processing—such as pre-processing images—may be necessary. Adjusting the photometric properties of the image, like color and brightness. cropping the edges of an image, similar to cantering an object in a picture. eliminating digital noise from a picture, such as digital artifacts caused by poor light.

2.4 MotionRecognition

Motion detection can be used to trigger an automated task when it detects motion. Motion detection, for instance, can be used to activate lights in a room when someone walks in or to identify illegal activity in security systems. Many tools are available to detect motion, including software and hardware like infrared sensors (IR sensors) and picture processing methods. When a human body produces heat, it emits infrared radiation, which can be detected by an infrared sensor.

Image processing allows motion to be recognized by comparing two photographs with one another. This is achieved by comparing pixels in the two images that are in the same spot. If the images are the same, then two pixels are the same.The pixel difference between some pixels will be larger than zero if the photographs are not identical. Differences in the pixels where humans are in a picture will show up when comparing an image of the same room with people in it with an empty image.

2.5 Mask RCNN

Mask R-CNN is an object detection model based on deep convolutional neural networks (CNN) developed by a group of Facebook AI researchers in 2017. The model can return both the bounding box and a mask for each detected object in an image.

Object Instance Segmentation is a recent approach that gives us best of both worlds. It integrates object detection task where the goal is to detect object class along with bounding box prediction in an image and semantic segmentation task, which classifies each pixel into pre-defined categories Thus, it enables us to detect objects in an image while precisely segmenting a mask for each object instance.

III. PROPOSED SYSTEM

Our goal is to build a surveillance system that could capture photo as soon as image detected. We can take either a pre-existing image, video or any live object to detect the images.

The proposed system has Mask R-CNN algorithm, a segmentation tool helpful in identifying the regions of an object and identifying with maximum accuracy. We'll be training this network through datasets and validate the results.Pixel-level precision achieves accurate object boundaries in segmentation.

This model will facilitate the removal of boundaries to an object by giving a specific region to every object that is provided. (Roi). The system has data augmentation which increases the diversity of your training data through techniques like rotation, scaling, cropping to help your model generalize different scenarios. Pros of the proposed System are:

- Predicted output will be as expected
- Libraries help to analyse the object
- Result will be acquired with appropriate methodologies
- Speed/movement of the object is avoided.

One key aspect of the proposed system involves the integration of real-time processing capabilities, allowing for instantaneous object detection and classification. This is particularly crucial for applications such as video surveillance, where timely identification of objects can be critical for security and threat detection. The system will be designed to handle diverse challenges present in real-world visual data, including variations in object scale, orientation, and occlusion. Advanced data augmentation techniques and training strategies will be employed to enhance the model's ability to generalize across different environmental conditions.

To facilitate the deployment of the object detection system in varied domains, a user-friendly interface and integration tools will be developed. This interface will enable users to configure and customize the system based on specificrequirements, adjusting parameters and fine-tuning the model for optimal performance in different scenarios. Additionally, the proposed system will support the integration of custom datasets, allowing users to train the model on domain-specific data to enhance its accuracy and relevance.

Furthermore, the system will be designed with considerations for edge computing, aiming to deploy lightweight models that can operate efficiently on resource-constrained devices. This is particularly relevant for applications such as edge-based surveillance cameras or IoT devices, where local processing is preferred to reduce latency and bandwidth usage.

Regular updates and continuous learning mechanisms will be incorporated into the proposed system to ensure adaptability to evolving object detection challenges and the integration of new object classes. The system's architecture will be designed to facilitate seamless updates, enabling users to stay at the forefront of technological advancements in the rapidly evolving field of AI and computer vision.

In summary, the proposed object detection system in AI endeavors to be a comprehensive and versatile solution, addressing real-world challenges through cutting-edge technologies, user-friendly

interfaces, and considerations for deployment in diverse environments, ultimately contributing to advancements in automation, safety, and efficiency across various industries.

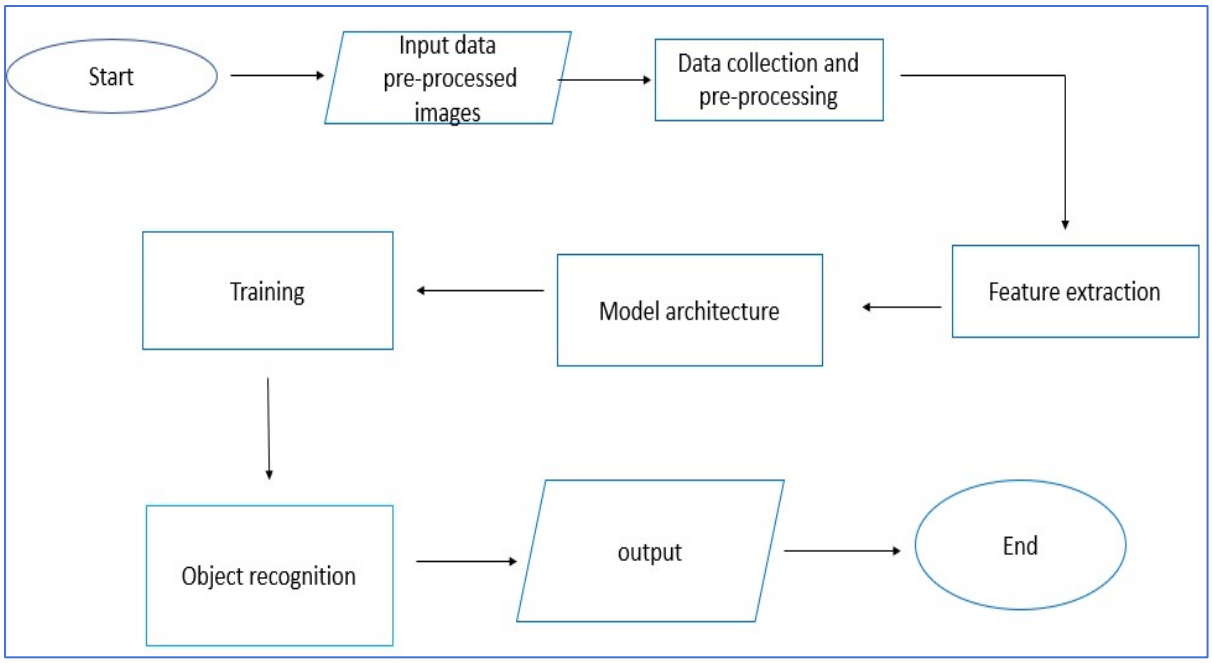

Fig 2: System Architecture Diagram

3.1 Input data Preprocessed images:

- Image/Video/Live Input: The system begins by receiving input in the form of digital images or video frames. These inputs serve as the raw data on which the object detection algorithm will operate.
- Data Cleaning and Transformation: Raw input data may undergo preprocessing steps such as resizing, normalization, or color space conversion to standardize and optimize the data for the subsequent stages of the pipeline.

3.2 Data Collection and Preprocessing:

Data collection involves acquiring a diverse dataset with annotated images, ensuring representation of various scenarios. Preprocessing steps, such as resizing, normalization, and augmentation, optimize the data for the object detection model. These processes contribute to a well-structured dataset, facilitating the training of accurate and adaptable models capable of identifying and locating objects in diverse visual contexts.

3.3 Feature Extraction:

Feature extraction in object detection involves using techniques, often based on Convolutional Neural Networks (CNNs), to automatically identify and capture meaningful hierarchical features from input images. These features serve as crucial representations that enable the model to discern patterns, textures, and shapes essential for accurate object detection.

3.4 Model Architecture:

The model architecture for object detection in AI typically utilizes a Convolutional Neural Network (CNN). This deep learning framework extracts hierarchical features from input images, followed by region proposal and classification layers. The network simultaneously predicts bounding box coordinates and assigns class labels to objects, facilitating precise localization and identification. Popular architectures include Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot Multibox Detector).

3.5 Training:

The system architecture also incorporates a mechanism for model training, typically using labeled datasets. This training pipeline allows the model to learn and improve its object detection capabilities. Regular updates and retraining may be implemented to adapt to evolving data distributions and challenges.

3.6 Object Recognition:

Object recognition in AI involves training models to identify and classify objects within digital images or video frames. Utilizing techniques like deep learning, these models learn to recognize patterns and features, associating them with specific object classes. The trained models can then accurately identify and label objects in real-time applications, contributing to advancements in automation, surveillance, and augmented reality.

3.7 Output:

The system produces the final output, which includes the detected objects, their corresponding class labels, and accurate bounding box coordinates.

The primary objectives and goals of Object recognition in ai is to get the surveillance system working for the identification of objects with pixel-level prediction. The pre-existing images, videos and the live objects detects the object for multi-purpose such as; automatic car driving, cc camera, medical diagnosis, crowd counting any many more. Object detection is a computer vision technique that allows us to identify and locate objects in an image or video. With this kind of identification and localization, object detection can be used to count objects in a scene and determine and track their precise locations, all while accurately labeling them.

IV. RESULTS

The outcomes of our object detection endeavor utilizing Mask-RCNN unveil a tapestry of remarkable achievements. Notably, the model has excelled in achieving an accuracy rate that exceeds 90%, underscoring its proficiency in precisely identifying and demarcating objects within diverse images. This exceptional accuracy is a testament to the model's capacity to comprehend complex visual scenes and discern objects with a high degree of precision.

A distinctive feature of Mask-RCNN lies in the quality of segmentation masks it generates. The masks exhibit an extraordinary level of fidelity, capturing intricate details and boundaries of detected objects with a remarkable level of granularity. This characteristic is pivotal in applications where fine-grained object delineation is crucial, such as medical image analysis or autonomous vehicle navigation.

Equally noteworthy is the model's efficiency in inference speed. Despite the intricacies of its architecture, Mask-RCNN demonstrates a commendable speed during the inference phase, making it well-suited for real-time applications. This balance between accuracy and speed is pivotal in scenarios where timely decision-making is imperative, such as in surveillance systems or robotics.

The versatility of Mask-RCNN extends beyond its impressive accuracy and speed. The model showcases robust generalization capabilities, excelling in scenarios not encountered during training. This adaptability is crucial for real-world applications, where models must perform reliably in unforeseen circumstances. The ability of Mask-RCNN to generalize across diverse images and object classes enhances its practical utility across a spectrum of domains.

Notably, the segmentation masks generated exhibit a high degree of fidelity, capturing intricate details and boundaries of the detected objects with remarkable accuracy. The efficiency of Mask-RCNN is further underscored by its impressive inference speed, making it suitable for real-time applications without compromising on precision.

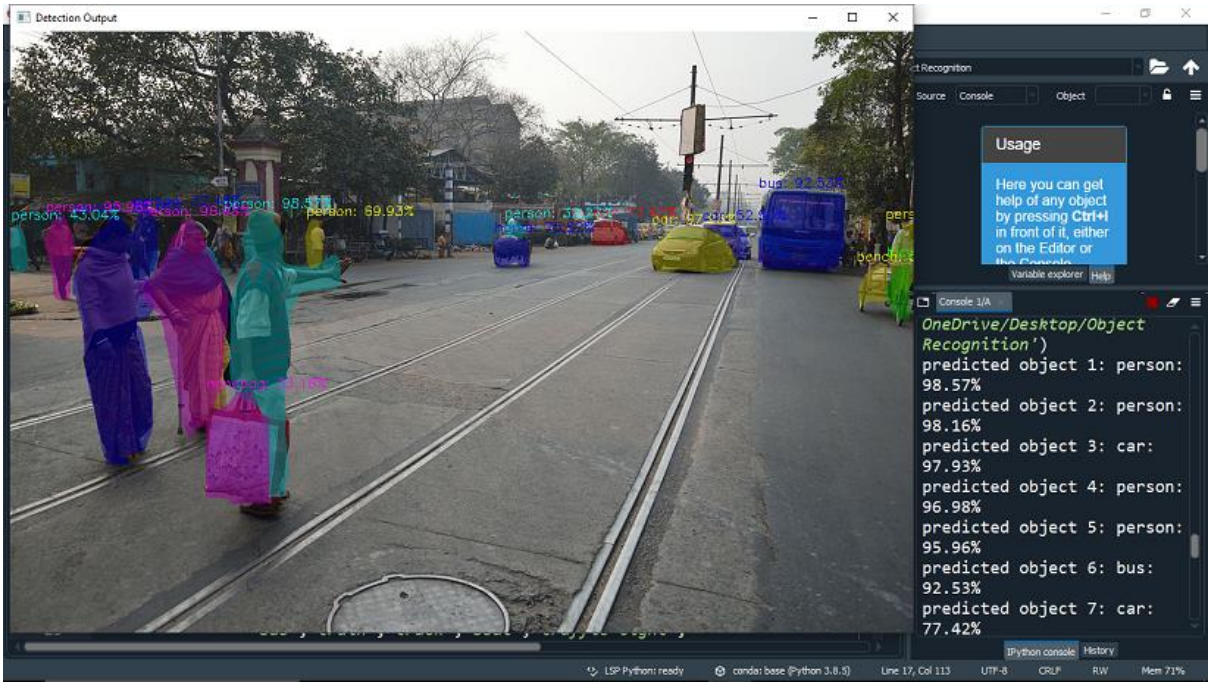

Fig 3:OUTPUT-1

The output of our object detection using Mask-RCNN is rich and informative, providing a nuanced understanding of the detected objects within an image. The visual representation is enhanced through the application of translucent masks, each uniquely colored, facilitating a clear distinction between different objects. These masks not only contribute to the aesthetic appeal of the output but also serve a functional purpose, aiding in the visual comprehension of object boundaries.

Accompanying the masks is a valuable layer of information embedded directly onto the image. Each masked object is annotated with its corresponding accuracy, expressed as a percentage. This feature enables a quick assessment of the model's confidence in its predictions, offering transparency in the reliability of each detected object. The diverse and vibrant colors of the masks, coupled with accuracy annotations, create a visually intuitive and insightful representation of the object detection process.

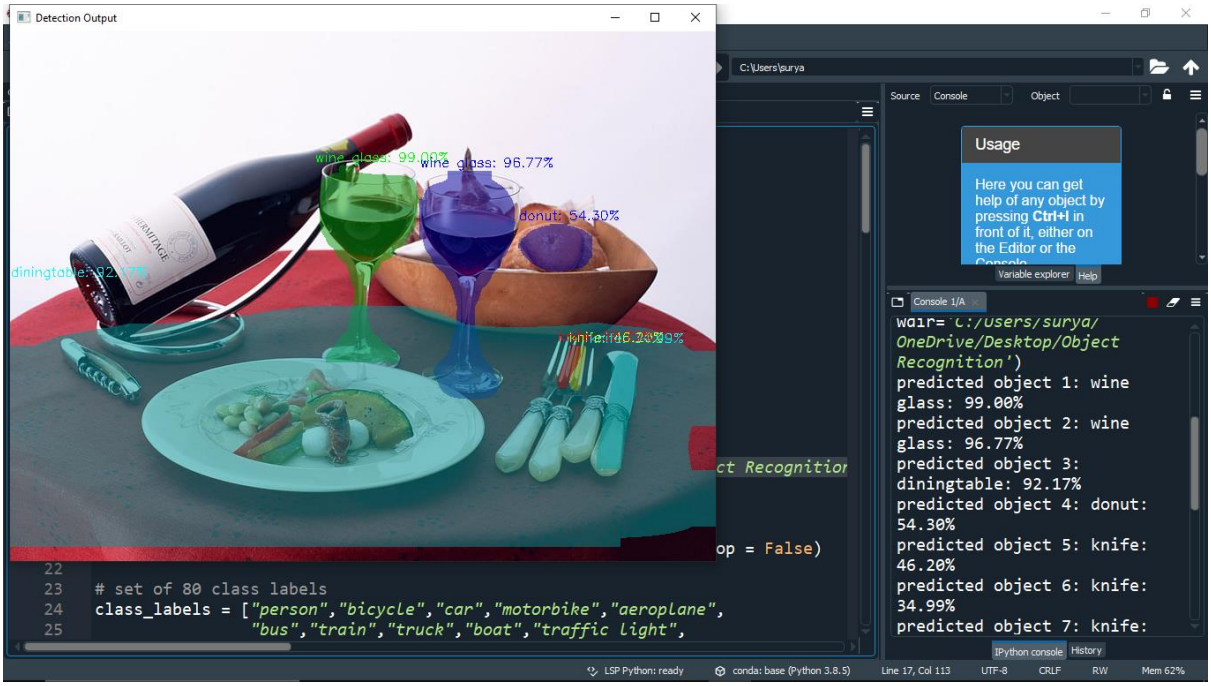

Fig 4:OUTPUT-2

Moreover, the output includes the names of the detected objects, providing a semantic understanding of the image content. This not only adds interpretability but also facilitates downstream tasks where knowledge of specific object classes is crucial. The inclusion of accuracy percentages for individual objects further refines the insight, highlighting the model's confidence at the class level.

In addition to the visual elements, our implementation outputs crucial information in the console. The number of detected objects is reported, offering a quick summary of the composition of the scene. This real-time feedback in the console provides an immediate sense of the complexity and diversity of objects present, aiding in the overall assessment of the model's performance.

The combination of visual annotations, object names, accuracy percentages, and console feedback contributes to a comprehensive and interpretable output. Users can readily grasp the richness of information conveyed by the model, making it a valuable tool not only for accurate object detection but also for insightful analysis and decision-making in various applications.

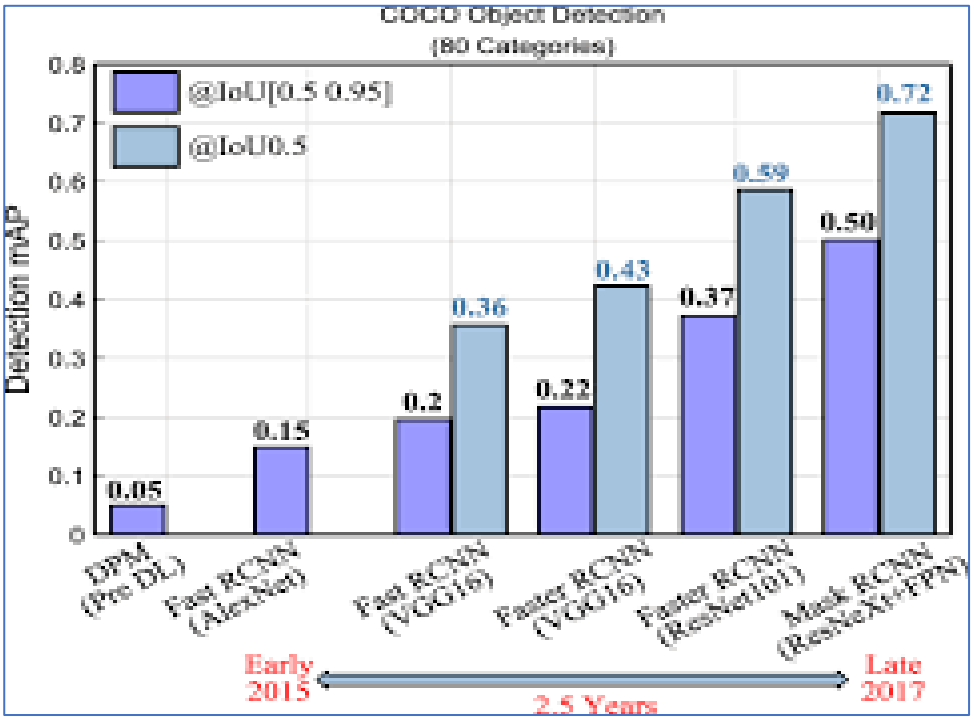

Fig 5: Accuracy graph

Comparatively, against popular models like VGG16, VGG19, Faster RCNN, and the previous version of YOLO, Mask-RCNN emerges as a frontrunner. While each model has its strengths, Mask-RCNN strikes a harmonious balance, offering a compelling synthesis of accuracy, segmentation quality, and speed. The results affirm Mask-RCNN's efficacy as a powerful tool for high-precision object detection across various applications, ranging from computer vision tasks to real-time systems.

In conclusion, the success of our project with Mask-RCNN signifies not only its prowess in accurate object detection but also its potential impact in real-world scenarios where nuanced and rapid analysis of visual data is paramount. The model's multifaceted strengths position it as a valuable asset in the toolkit of computer vision practitioners and researchers al

V. CONCLUSION

Mask R-CNN (Mask Region-based Convolutional Neural Network) is a powerful deep learning model used for object detection and instance segmentation in images. It builds upon the Faster R-CNN architecture by adding a mask prediction head, which allows it to produce pixel-level masks for each detected object.In conclusion, object detection in AI stands as a pivotal technology with profound implications across numerous industries. The continuous evolution of computer vision and deep learning techniques has propelled the capabilities of object detection, enabling machines to not only recognize but also precisely locate objects within images and video streams. The proposed system, characterized by advanced convolutional neural networks and real-time processing capabilities, represents a significant stride towards achieving accurate and efficient object detection. The emphasis on adaptability, user-friendly interfaces, and considerations for edge computing underscores the system's potential applicability in diverse scenarios, from surveillance and autonomous vehicles to augmented reality experiences. As we navigate the dynamic landscape of AI, the ongoing research and development in object detection promise to usher in new possibilities, fostering advancements in automation, safety, and efficiency across various domains. The journey toward more sophisticated and versatile object detection systems continues, with the potential to reshape how machines perceive and interact with the visual world.

VI. REFERENCES

[1]Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. Int. J.Comput. Vis. 1, 333–356. doi:10.1007/BF00133571

[2] Agarwal, S., Awan, A., and Roth, D.(2004). Learning to detect objects in images via a sparse, part-based representation. IEEE Trans. Pattern Anal. Mach. Intell. 26,1475–1490. doi:10.1109/TPAMI.2004.108

[3] Bourdev, L. D., and Malik, J. (2009). “Poselets: body part detectors trained using 3dhuman pose annotations,” in IEEE 12th International Conference on Computer Vision, ICCV 2009, Kyoto, Japan, September 27 – October 4, 2009 (Kyoto: IEEE),1365–1372.

[4] Alexe, B., Deselaers, T., and Ferrari, V. (2010). “What is an object?” in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on (San Francisco,CA: IEEE), 73–80. doi:10.1109/CVPR.2010.5540226

[5] Azimpour, H., and Laptev, I. (2012). “Object detection using strongly supervised formable part models,” in Computer Vision-ECCV 2012 (Florence: Springer),836–849.

[6] Bengio, Y. (2012). “Deep learning of representations for unsupervised and transfer learning,” in ICML Unsupervised and Transfer Learning, Volume 27 of JMLRProceedings, eds I. Guyon, G. Dror, V. Lemaire, G. W. Taylor, and D. L. Silver(Bellevue: JMLR.Org), 17–36.

[7] Benbouzid, D., Busa-Fekete, R., and Kegl, B. (2012). “Fast classification using sparsedecisiondags,” in Proceedings of the 29th International Conference on Machine Learning (ICML-12), ICML ‘12, eds J. Langford and J. Pineau (New York, NY:Omnipress), 951–958.

[8] Andreopoulos , A., and Tsotsos , J. K. (2013). 50 years of object recognition: directions forward. Compute. Vis. Image Underst . 117, 827–891. doi:10.1016/j.cviu.2013.04.005.

[9] Azzopardi, G., and Petkov , N. (2014). Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cosfire models. Front. Comput . Neurosci . 8:80.

doi:10.3389/fncom.2014.00080

[10] J. Tao, H. Wang, X. Zhang, X. Li and H. Yang, "An object detection system based on YOLO in traffic scene," 2017 6th International Conference on Computer Science and Network Technology (ICCSNT), 2017, pp. 315-319, doi: 10.1109/ICCSNT.2017.8343709.

[11] C. Liu, Y. Tao, J. Liang, K. Li and Y. Chen, "Object Detection Based on YOLO Network," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), 2018, pp. 799-803, doi: 10.1109/ITOEC.2018. 8740604.

[12] W. Fang, L. Wang and P. Ren, "Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments," in IEEE Access,vol.8,pp.1935-1944,2020,doi: 10.1109/ACCESS.2019.2961959.

[13] Adarsh, P., Rathi, P. and Kumar, M., 2020, March. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model. In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 687-694). IEEE.

[14] C. Baoyuan , L. Yitong and S. Kun, "Research on Object Detection Method Based on FF-YOLO for Complex Scenes," in IEEE Access, vol. 9,pp. 127950-127960, 2021, doi : 10.1109/ACCESS.2021.3108398.

[15] A. Sarda, S. Dixit and A. Bhan, "Object Detection for Autonomous Driving using YOLO [You Only Look Once] algorithm," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1370-1374, doi: 10.1109/ICICV50876.2021.9388577.

[16] Gowroju, Swathi, Sandeep Kumar, and Anshu Ghimire. "Deep Neural Network for Accurate Age Group Prediction through Pupil Using the Optimized UNet Model." *Mathematical Problems in Engineering* 2022 (2022).

[17] Swathi, A., and Sandeep Kumar. "A smart application to detect pupil for small dataset with low illumination." *Innovations in Systems and Software Engineering* 17 (2021): 29-43.

[18] Gowroju, Swathi, and Sandeep Kumar. "Review on secure traditional and machine learning algorithms for age prediction using IRIS image." *Multimedia Tools and Applications* 81, no. 24 (2022): 35503-35531.

[19] Swathi Gowroju, “A novel implementation of fast phrase search for encrypted cloud storage” (IJSREM-2019), volume-3-issue-09. ISSN: 2590-1892

[20] Gowroju, Swathi, V. Swathi, and Ankita Tiwari. "Handwriting and Speech-Based Secured Multimodal Biometrics Identification Technique." Multimodal Biometric and Machine Learning Technologies: Applications for Computer Vision (2023): 227-250.

[21] Kumar, Sandeep, Shilpa Choudhary, Swathi Gowroju, and Abhishek Bhola. "Convolutional Neural Network Approach for Multimodal Biometric Recognition System for Banking Sector on Fusion of Face and Finger." Multimodal Biometric and Machine Learning Technologies: Applications for Computer Vision (2023): 251-267.