

Real Time Object Detection Tracking Using YOLO and Deep Sort

Hemashree C¹, Pallavi B S², Pruthvi H N³, Shivaganga M⁴, Dr. Santhosh S⁵

^{1,2,3,4} B.E. Scholars, Department of CSE

⁵Professor, Department of CSE

Kalpataru Institute of Technology, Tiptur, India

Abstract: For real-time object detection and speech alarms, Research offers a sophisticated blind aid solution that smoothly combines the YOLO algorithm with Open CV's DNN (Deep Neural Network) module. The main objective is to improve the safety and freedom of people with visual impairments by offering quick object recognition and audio feedback. The YOLO method, which is tuned for real-time inference, is used by the system to precisely recognize objects after using a webcam to record live video input. Then, using audio description, it produces speech notifications that provide crucial details about the things it has spotted. Because of its exceptional versatility, the research can provide speech outputs in the user's preferred language, increasing its usability and accessibility. Its versatility is further demonstrated by its capacity to precisely handle a variety of object classes, making it a priceless tool for greatly enhancing the lives of those who are blind or visually impaired. Challenges in real-time object detection include occlusion, scale variations, and cluttered environments. Researchers must navigate the trade-offs between accuracy and speed. Real-time object detection is pivotal in computer vision, enabling intelligent systems across diverse applications. Real-time object detection systems designed for assistive technologies are becoming increasingly important in empowering individuals with visual impairments. The research integrates the YOLO algorithm with OpenCV's DNN module to deliver a highly responsive blind-aid solution capable of capturing live video through a webcam and processing it instantly. YOLO's architecture, optimized for speed and accuracy, enables the system to detect and classify multiple objects within each video frame, even in dynamic or unpredictable environments. Once an object is identified, the system generates immediate speech output, providing descriptive audio alerts in the user's preferred language, thus enhancing accessibility and user experience. This multilingual capability makes the solution versatile across different regions and cultures. Additionally, the system supports a wide range of object categories such as vehicles, household items, and obstacles helping visually impaired users navigate daily surroundings more safely and independently. However, the implementation faces challenges, including handling occlusions, varying object scales, low-light conditions, and cluttered backgrounds. These factors require careful balancing between detection accuracy and processing speed to maintain real-time performance. Despite these challenges, the integration of advanced deep learning models with user-friendly audio interfaces demonstrates significant potential to improve mobility, safety, and overall quality of life for people with visual impairments.

Keywords: Artificial Intelligence and Computer vision, YOLO, Deep SORT Algorithm

