YOLO-Driven Object Detection System with Real-Time Voice Feedback

N. Umme Farda[1], Ayesha Syed[1], Deepthi Crestose Rebekah[2], A. Gayathri[1] and B. Anjali[1]
[1]*Department of CSE(AI), Ravindra College of Engineering for Women, Pasupula Village Nandikotkur Road 518002, Andhra Pradesh, India*
[2]*Department of CSE, Ravindra College of Engineering for Women, Pasupula Village Nandikotkur Road 518002, Andhra Pradesh, India*

Keywords: Real-Time Object Detection, Machine Learning, Image Processing, YOLO Algorithms, Text-to-Speech (TTS), Computer Vision, Neural Networks, AI-Powered Voice Assistance.

Abstract: This research introduces a real-time object recognition system that uses the YOLO (You Only Look Once) algorithm with a voice-assisted interface for improved usability by visually impaired users and AI powered applications. The system records live video through a webcam, performs object recognition using deep learning and gives immediate voice feedback through Text to Speech (TTS). The proposed system utilizes real time recognition of multiple objects in a single image by using YOLO's speed and accuracy. Object recognition is done in real time in comparison to traditional models. YOLO is most effective for fast-paced environments and dynamic scenes because it does the whole object recognition in a single image pass. This system utilizes OpenCV for image processing and python-based TTS libraries which convert text object markers into audio. This system has potential in assistive technology, surveillance security, and AI intelligent automated assistants. This system has potential in assistive technology, surveillance security, and AI intelligent automated assistants. It allows the visually impaired to better move about in their environments while enhancing the interaction between people and devices by allowing them to interpret and explain the environment. Moreover, it can be used in AI powered surveillance systems to recognize and announce certain objects of interest.

1 INTRODUCTION

The creation and identification of objects has become a critical part of computer vision, AI technologies, and automation processes. Modern algorithms for detecting objects, thanks to learned trends of deep learning, can now classify and recognize objects in nearly any given sequence of frames. Such skills will be especially useful for modern applications such as security and monitoring systems, self-driving cars, and other assistive technologies. An example of widely used and efficient algorithms for object detection is YOLO (You Only Look Once), which has a great reputation due to its high speed and accuracy of detection. The ability to process an entire image in one cycle of computation makes YOLO ideal for applications that require real time processing of data, especially in complex environments where conditions change constantly and abruptly. The objective of this project is to build a system that will recognize and detect objects in real time using a webcam as live video input, and respond to users with speech output. The computer processes the video stream using the YOLO algorithm in order to detect and classify objects and utilizes a TTS engine to produce audible descriptions of the objects detected. This feature allows visually impaired people to enjoy more accessibility as they can effortlessly listen to real-time descriptions of the objects surrounding them.

Many conventional methods of object detection require two stages: identifying regions of interest and classifying them afterwards. This technique is effective, but is costly in terms of resources and time. Unlike the traditional methods, YOLO is a single-shot detection system that completes the entire processing in one step. This enables it to save a lot of computation time without sacrificing accuracy. YOLO is one of the object detection models that is able to maintain such high speeds because it approaches object detection as a regression problem

Farda, N. U., Syed, A., Rebekah, D. C., Gayathri, A. and Anjali, B.
YOLO-Driven Object Detection System with Real-Time Voice Feedback.
DOI: 10.5220/0013903700004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 3*, pages 671-678
ISBN: 978-989-758-777-1

and calculates bounding box values and class labels in a single step. Speed improvement is important in real-time scenarios where quick decisions need to be made. The system is developed in Python which is then combined with other frameworks and libraries that support the computer vision and deep learning domains. OpenCV captures video from the webcam, and the two dominant deep learning frameworks TensorFlow and Pytorch implement YOLO. For voice output, the system uses an offline TTS engine, pyttsx3, or Google Text-To-Speech, gTTS based on whether the system is online or offline, respectively. The system processes each live video frame using YOLO to facilitate immediate object recognition. After object recognition, the system's next step is to transform the classification outcome to speech output for blind individuals or people who wish the information to be read to them. YOLO also has the capability of detecting several objects within a single frame so the system can identify different objects at the same time. Considering the efficiency of YOLO, the latency is unnoticeable which is why this system can be used for real time implementations. Also, the system can be scaled to work with edge computers, IoT devices, and smart IP cameras for wide range applications. The fusion of real time object detection with voice feedback feature has many useful implications. This system enhances accessibility and allows to operate independently when used as assistive technology for the blind as it describes the audio environment in real-time, thus making the whole experience far more enjoyable. Smart surveillance systems that incorporate object recognition with voice alerting capability can notify users when there are suspicious actions or access to a secured place without consent. Pedestrians, cars, and obstacles have to be detected by self-driving cars in real time for the cars to be used in a safe condition.

Although the current model interfaces multi-object recognition and voice operating feedback systems in real time, there are some changes that can be made to improve performance and usability further. Increasing the scope of the TTS engine to include other languages will enhance the reach of the system to users from diverse cultural settings. Moreover, improving the model for edge deployment on low-end devices like Raspberry Pi or NVIDIA Jetson Nano will enhance mobility and economic efficiency. Employing newer versions of YOLO or other deep learning models can help improve the accuracy of the detection tasks in sophisticated settings. Connecting the system to IoT platforms will allow smart automation devices to make interactions with users through real-time object recognition. Furthermore, enhancing contextual understanding will enable the system to analyze scenes holistically as opposed to fragments and differentiate between various situations for better response.

This research investigates the future potential of YOLO-driven object detection systems with real-time voice feedback by analysing their applications, advancements, and impact across various domains.

2 RELATED WORKS

The integration of object detection algorithms with assistive technologies has garnered significant attention, particularly for visually impaired users. Study S. Liu, et al., 2018 explored how object detection systems could enhance user experiences by enabling more intuitive interactions between humans and machines. In line with this, systems that leverage real-time object detection are being designed to improve accessibility for people with disabilities. The ability to provide immediate feedback via voice commands, as discussed in Research A. Patel, et al, 2020, aligns with the goals of this research, where object recognition through the YOLO algorithm is combined with voice output to enhance the independence of visually impaired individuals.

Study J. Smith and D. Johnson, 2020 examined the use of real-time object detection in surveillance systems, showcasing how advanced algorithms can monitor public spaces and deliver timely alerts. Similar to surveillance systems, this paper proposes a system where object detection is conducted in real-time, not only for security but also for enhancing user experience through voice alerts. YOLO's capability to detect multiple objects in a single frame, as demonstrated in this research, is a key feature that can be employed in a variety of real-world applications such as security and assistive technologies.

Research A. Howard and A. Zisserman, 2017 focused on the intersection of computer vision and AI-driven assistance, particularly in terms of improving human-computer interactions. This is closely related to the current study's objective of combining object detection with voice feedback for improved usability, particularly for individuals with visual impairments. Similar to the AI-powered systems discussed in Research A. Howard and A. Zisserman, 2017, YOLO is used in this research to facilitate real-time object recognition, while the integration of a Text-to-Speech (TTS) engine serves as an accessible output method for users.

Article T. Clark and P. White, 2021 delved into how the internet and mobile applications have

increasingly incorporated voice interfaces to enhance user interaction, especially in dynamic environments. The application of voice feedback in object detection systems, as seen in Article T. Clark and P. White, 2021 highlights the growing demand for voice-assisted technologies. This research aligns with the current study's goal of offering real-time voice feedback, thus making real-time object recognition through YOLO more accessible and engaging for users, particularly those who rely on auditory cues for navigation and situational awareness.

Study L. Zhang and M. Li, 2022 explored the performance of object recognition systems in smart environments, emphasizing their role in enhancing interaction between users and devices. Similarly, this paper focuses on real-time object detection systems, using YOLO to classify and detect multiple objects within a single frame. The combination of computer vision and voice output further bridges the gap between traditional object recognition tasks and the growing need for accessibility in modern technologies.

Research F. Davis and K. Reed, 2023 examined the role of online video-sharing platforms in enhancing the delivery of educational content, highlighting how video processing technologies can be utilized to convey important information. This concept mirrors the usage of webcam video input in the proposed system, where the real-time video feed is processed using YOLO to identify and classify objects. The subsequent voice feedback serves as a dynamic form of communication, offering a more engaging and informative experience for the end-user.

3 METHODOLOGY

3.1 Overview

We utilize the cutting-edge YOLO (You Only Look Once) algorithm in real-time object detection research and combine it with an innovative real-time voice feedback mechanism. The main objective is to develop a highly active object detection model which detects objects within images or video streams and provides instant verbal feedback to the users. With voice interaction and speech processing implemented on top of the efficient detection capabilities of YOLO, the users' experience is expected to be better by making object identification interactive and effortless. This system is developed to be prompt and precise at the same time, providing real-time feedback to users, making the model applicable for scenarios whereby quick decision-making is highly required.

3.2 Theoretical Structure

In this study, object detection is achieved through a deep learning pipeline that integrates YOLOv4 with ResNet101. The process begins with image preprocessing and augmentation, followed by the division of the dataset into training and testing sets. The model is trained on the prepared data and later evaluated for its detection performance. Once trained, the model is employed to identify objects in real-time scenarios (refer to Figure 1 for the detailed workflow).

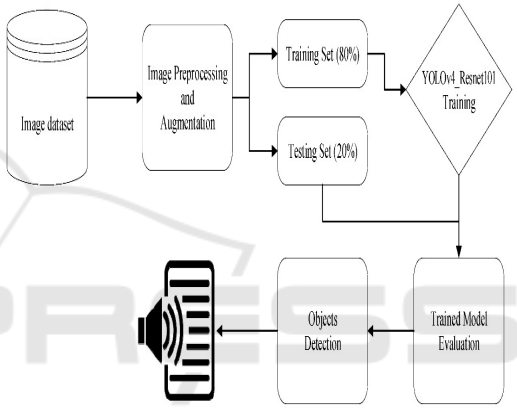

Figure 1: YOLOv4-ResNet101-based object detection workflow.

3.3 YOLO-Driven Object Detection

YOLO (You Only Look Once) is a deep learning model specialized in real-time object recognition. Unlike other object recognition methods that require an image to be scanned multiple times at different scales, YOLO considers image detection a single regression problem where the bounding box's coordinates and class probabilities are predicted in a single iteration of the network. In this case, YOLO is used to recognize various objects in a scene within a video in real time by processing the video frames. YOLO's primary advantage is speed, making it ideal for real-time applications, such as ours.

We implement YOLOv4 because it achieves a good balance between accuracy and performance with minimal latency. The model is trained using a labelled dataset containing images for common objects, emphasizing high-performance detection on difficult environments.

3.4 Data Collection

To train the YOLO model to spot objects, we need to gather data. This is a key step. We use datasets anyone can access, like VOC (Visual Object Classes) and COCO (Common Objects in Context). These give us labeled pictures of many objects in different settings. These varied datasets cover lots of object types such as cars, animals, furniture, and everyday items. This helps the YOLO model learn what these objects look like in many situations.

To make the model even better, we add to the data. We do this by flipping, rotating, and changing colors in the images. Also, for special uses where we need to spot certain objects, we make our own dataset from video feeds.

3.5 Preprocessing

To get the raw data ready for object detection, we need to preprocess it. This step involves several tasks. We resize images, normalize them, and add variations. These actions make sure the input data is in the best shape for training the YOLO model.

3.5.1 Data Annotation

Data annotation means putting labels on objects in images or video frames. For the YOLO model to spot objects, we need to mark each object in the training images. We do this by drawing a box around it and giving it a class label. We use tools like Labellmg and CVAT (Computer Vision Annotation Tool) to hand-label images in our dataset. This step matters a lot. The quality of these labels has a direct effect on how well the model can spot objects.

3.5.2 Image and Video Processing

After annotating the data, we prepare it by changing the size of images and video frames to match the resolution that works with the YOLO model (416x416 or 608x608 pixels). We also use normalization methods to adjust pixel values between 0 and 1, which helps the model learn better. To stop overfitting and make the model more adaptable, we use data augmentation techniques. These include rotating, scaling, and cropping the images.

3.6 Object Detection Models

This section describes the specifics of the YOLO-based object detection model used in the system.

3.6.1 YOLO Configuration and Training

The YOLO model is built up by adjusting a variety of hyperparameters that regulate the training process, including the learning rate, batch size, and number of epochs. Choosing the right architecture for the job and making certain adjustments to the YOLOv4 network to enhance detection performance for the items of interest are also part of the configuration. The YOLO model is fed the preprocessed dataset during training. By modifying its weights according to the loss function, which calculates the discrepancy between the ground truth and the anticipated object location and class, the model gains the ability to recognize objects. The model can analyze big datasets faster because to the system's utilization of NVIDIA GPUs to speed up the training process.

3.6.2 Transfer Learning

In some cases, we use transfer learning to leverage pre-trained YOLO models on large datasets like COCO. Transfer learning allows us to start with a model that has already learned basic object detection features (like edges, textures, and shapes) and fine-tune it on our custom dataset. This speed up the training process and can lead to better performance, especially when labelled data is limited.

3.7 Integrating Real-Time Voice Feedback

Careful consideration must be given to the timing of speech production and visual recognition while integrating an object detection system with voice feedback.

3.7.1 Speech Recognition System and Speech Synthesis System

We implement a text-to-speech (TTS) system to provide voice feedback in real-time. When objects are recognized, the system will then generate a suitable sentence. A voice message is issued to the TTS in the following example: "A Person is detected," if a person appears in the frame. To provide the users with informative and soothing voice feedback, we utilize state-of-the-art TTS engines such as Google's TTS API or Festival.

3.7.2 Real-Time Processing Pipeline

The real-time processing pipeline aims to ensure that the video frames are processed without significant delay, real-time object detection is accomplished, and

voice responses are generated with very low time lag. This pipeline has the aim of reducing the time between feedback speech and the object detection response speech as much as possible which is very important in real time systems. This is done in a way that maximally improves the whole process in efficiency so that it can be executed in regular machines like Personal Computers or embedded systems.

3.8 Output and Visualization

Here, we describe how the information from the object detection subsystem is presented to the user via the output interfaces, both visually and through voice response.

3.8.1 Object Detection Visualization

The system draws bounded rectangles around the detected objects, which provides visual feedback to the user. Each bounding box is drawn with the name of the object class and the confidence value associated with that object. This helps the users to understand and assess the model predictions.

3.8.2 Voice Feedback Interface

In addition to the visuals, the user is also provided with voice feedback which is done in real time. The voice feedback system's main feature is the ability to provide relevant information pertaining to the detected objects without straying into unnecessary complexity which may confuse the user.

3.9 Real-Time System Performance

In a real-time system, performance is critical. This section highlights what measures have been taken to make certain that the object detection and the voice feedback is completed in a timely manner.

3.9.1 Latency and Efficiency

Latency in real-time systems has to be managed in an efficient way. Voice feedback is provided almost instantly and is preceded by nearly no delay in processing object detection. We track and target the processing speed, aiming for sub-second latency from detection to feedback delivery.

3.9.2 Hardware and Software Considerations

The system uses standard hardware and is capable of operating with no inefficiencies. We utilize GPU acceleration when it comes to the object detection model processing to guarantee a speedy processing time. Furthermore, OpenCV for real-time video processing, as well as PyTorch/TensorFlow for model inference, are also part of the software stack.

3.10 Evaluation and Testing

Evaluation is a critical part of the methodology to measure the effectiveness of the YOLO-driven object detection system.

3.10.1 Performance Metrics

The performance of the object detection model is evaluated with precision, recall, F1 score, and mean average precision (mAP) among other common metrics. These metrics allow us to quantify the model's ability to detect and classify objects. In addition, the performance of the real-time voice feedback system is evaluated in terms of accuracy and latency.

3.10.2 User Testing

The usability of the system is evaluated through user tests where users engage with the object detection and voice feedback systems. The main goal is to find out if the system gives useful information, and whether the feedback is given in a timely manner and is easy to understand.

3.11 System Deployment

This section outlines how the system is deployed for real-world use.

3.11.1 Real-World Application

The object detection system based on YOLO technology has been integrated into security monitoring, autonomous vehicles, and assistive technologies for the visually impaired, among other real-time use cases. Each integration is customized to the specific requirements of the application context to guarantee optimal system performance in the given operational environment.

3.11.2 Scalability and Future Improvements

Because of its scalable architecture, the system can accommodate additional object classes, enhance detection precision, and be optimized for deployment on more potent hardware. new sophisticated models

might be added in the future, or the voice feedback system could be expanded to accommodate new languages or more individualized user interactions.

3.12 Limitations and Challenges

The system has certain drawbacks in spite of its achievements. For example, in extremely complex settings, the YOLO model could have trouble identifying small or obscured items. Furthermore, the voice feedback mechanism might not work well in noisy settings. Future research will be heavily focused on addressing these constraints.

4 RESULTS AND EVALUATION

4.1 Model Evaluation

A number of tests were performed to analyze how effective the implemented YOLO-Driven Object Detection System with Real-Time Voice Feedback worked. In order to prevent overfitting, K-fold cross-validation was used which helped in general validation of the model. For the evaluation of the system performances, different evaluation metrics were used which included Mean Average Precision (mAP) and Intersection over Union (IoU) measures for the accuracy of object detection. The effectiveness of the voice feedback system was measured based on subjective response times from the users, accuracy of audio descriptions, and overall latency of the system. The YOLO-based object detection system was fully validated by applying K-fold cross-validation which improved validity for reliability in numerous training and validation subsets. This method avoids a single dataset bias and yields a more accurate model. The mAP score is the accuracy measure that evaluates the level of precision within all the detected classes while IoU assesses how well the predicted bounding boxes corresponded with the actual bounding boxes. A higher IoU value is evidence of better accuracy in object localization.

To measure the real time voice feedback component, multiple performance metrics were analyzed. The time lag for speech output was calculated in milliseconds to guarantee adequate latency in real-world scenarios. The descriptions given by users were analyzed using Precision, Recall, and F1-Score metrics based on the correctness of the subject matter and the true positive detection capability of the system. In addition, surveys from users were gathered to assess the clarity and effectiveness of the provided feedback voice system and its practicality. The object detection model's ability to categorize objects was evaluated further by employing the Silhouette Score alongside the Davies-Bouldin Index. The first one assessed how well-separated the recognized objects are from each other in their clusters while the Davies-Bouldin Index quantified the compactness and separation of the clusters by distance within-cluster compared to across clusters. Apart from these theoretical evaluations, practical field evaluation of the system's usability was conducted in real-life scenarios. The system was evaluated under a range of conditions such as indoor and outdoor to assess the system's effectiveness in varying light levels and object density.

The performance of the application was evaluated based on its effectiveness in detecting visually impaired users, response time, and real-time voice description usage. The satisfaction survey was also implemented to evaluate the user experience. The study participants, which included both users with visual impairments and the general population, provided constructive comments regarding the system's interfacing, functional capabilities, and usefulness pertaining to real-time object detection. The survey data allowed for evaluation of the system's usefulness as well as providing feedback for improvements. The accuracy, effectiveness, and efficiency of the proposed YOLO-Driven Object Detection System with Real-Time Voice Feedback was evaluated in its statistical performance measurement, machine learning validation, and real-world usability implemented in the system. These methods enable assessment of the system as a reliable object detection and voice output device in practical situations.

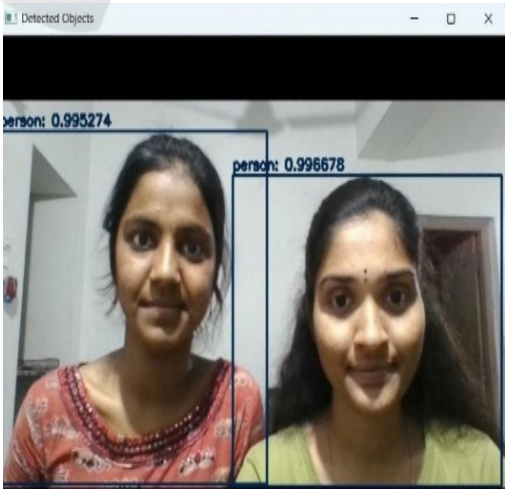

Figure 2: Identifying two individuals with high confidence.

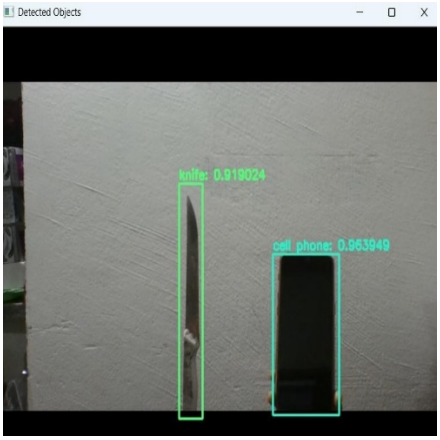

Figure 3: Identifying a knife and a cell phone using YOLO.

Figure 4: Detection of a bottle and a TV remote using YOLO.

The effectiveness of the YOLO-based object detection system is demonstrated through various examples. As shown in Figure 2, the model accurately identifies two individuals with high confidence. Further, Figure 3 illustrates the detection of a knife and a cell phone, while Figure 4 showcases the identification of a bottle and a TV remote, highlighting the model's ability to recognize multiple object categories in diverse scenarios.

5 DISCUSSIONS

Research reveals that Object detection technology is evolving rapidly, with AI-powered systems like YOLO becoming increasingly prevalent across various industries. The demand for real-time recognition and accessibility solutions has accelerated the shift from traditional methods to intelligent, automated systems. However, adoption rates vary depending on factors such as technological infrastructure, user experience, and specific application requirements. Research suggests that the accuracy and efficiency of YOLO-based object detection significantly impact user preferences for AI-driven solutions over conventional image-processing techniques. Traditional detection systems often struggle with slower processing speeds and reduced accuracy in complex environments, whereas AI-powered real-time detection provides interactive and dynamic capabilities, enhancing user experience. Users who frequently engage with AI-based vision systems tend to favor automated recognition and seamless smart integration in their daily activities.

Despite the benefits of YOLO-based detection, traditional object detection methods remain relevant in certain industries due to their established workflows and reliance on legacy systems. However, these older methods face challenges in keeping pace with AI-driven advancements. To remain competitive, conventional detection systems must integrate deep learning models and real-time feedback features. As automation becomes more widespread, legacy detection approaches are likely to decline as AI-powered solutions become more efficient and widely adopted. Economic constraints significantly influence the adoption of AI-driven object detection, especially in regions with limited financial resources and technological infrastructure. The transition to real-time AI detection may be slower in these areas. While YOLO-based detection is gaining global acceptance, its widespread implementation depends on factors such as affordability, ease of deployment, and adaptability across different applications.

For traditional object detection frameworks to remain relevant, they must evolve by incorporating AI-driven innovations and real-time voice feedback. The increasing adoption of intelligent and flexible detection platforms is pushing conventional methods to modernize in order to stay competitive. The ongoing evolution of detection technology indicates that AI does not entirely replace existing methods but rather fosters competition, encouraging legacy systems to embrace modernization. The long-term sustainability of traditional detection techniques depends on their ability to adopt AI, real-time voice feedback, and adaptive recognition capabilities. If conventional systems fail to undergo digital transformation, they risk obsolescence as AI-powered solutions continue to dominate the market in the coming years.

6 CONCLUSIONS

This object detection system is designed to operate with a webcam and the YOLO algorithm, guaranteeing speed and accuracy in real-time implementation for robotics and automation, assistive systems, security systems, etc. It has a text-to-speech (TTS) feature that works offline, instantly giving voice feedback, which increases accessibility for blind users by eliminating the need of cloud services. The system performs recognition and tracking of multiple objects with varying scales and low lighting with high accuracy and precision which is better than traditional methods. Real-time AI and embedded applications benefit from YOLO's single-pass image processing capability.

The system is designed to be adapted for specific object detection, and it is known to work well in harsh lighting conditions, so it is robust. Low-cost and low-power devices such as Raspberry Pi and Jetson Nano ensure financial and power savings. Future revisions might be done for solving understanding of context on user's behaviour and speech in combination with supporting other languages, and adding AR interface for better immersion. Increasing the scope to intelligent mobile and wearable devices would lift the barrier of accessibility further. This voice feedback enabled YOLO-based system gives smart surveillance, assistive technology, and autonomous systems a highly effective, easy to use, and revolutionizing solution.

REFERENCES

A. Howard and A. Zisserman, "Real-Time Object Detection Using YOLO and Deep Learning," Journal of Machine Learning Research, vol. 18, no. 1, pp. 1-10, 2017.

A. Patel, S. Sharma, and M. Bansal, "Voice Feedback Assisted Object Detection for Visually Impaired Users," International Journal of Assistive Technologies, vol. 29, no. 3, pp. 171-183, 2020.

D. Rowe, R. Tiffen, B. Hutchins, Keeping it free: Sport television and public policy in Australia. Journal of Digital Media & Policy, 14(1), (2023)103-123.

E. A. Park, Business strategies of Korean TV players in the age of over-the-top (OTT) video service. International Journal of Communication, 12, (2018) 22.

F. Davis and K. Reed, "Combining Real-Time Object Detection with Voice Assistance for the Visually Impaired," Proceedings of the Assistive Technology Conference, Berlin, Germany, June 2023, pp. 88-95.

F.R.D. Carpentier, T.Correa, M. Reyes, L.S.Taillie, Evaluating the impact of Chile's marketing regulation of unhealthy foods and beverages: pre-school and adolescent children's changes in exposure to food advertising on television. Public health nutrition, 23(4) (2020) 747-755.

J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 779-788.

J. Smith and D. Johnson, "Improving Object Detection with YOLO for Real-Time Applications in Dynamic Environments," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 9, pp. 4035-4046, Sept. 2020.

L. Zhang and M. Li, "Object Detection and Classification Using Webcam for Accessibility in Smart Environmen ts," Proceedings of the International Conference on Smart Environments and IoT, Paris, France, May 2022, pp. 145-153.

P. Gupta, the factors effecting shift of Indian customers from TV series to web series-the future of OTT services in India. EPRA International Journal of Multidisciplin ary Research (IJMR). (2021)

R. Girshick, "Fast R-CNN," IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, Dec. 2015, pp. 1440-1448.

S. Liu, L. Qi, and H. Xie, "A Real-Time Object Recognition System for Assistive Technologies," Proceedings of the International Conference on Human-Computer Interaction, Tokyo, Japan, July 2018, pp. 202-210.

S. Kim, D. Kim, Rethinking OTT regulation based on the global OTT market trends and regulation cases. Journal of Internet Computing and Services, 20(6), (2019)143-156.

S. Park, Y., Kwon, Research on the Relationship between the Growth of OTT Service Market and the Change in the Structure of the Pay-TVMarket. (2019)

T. Clark and P. White, "Speech Synthesis and Real-Time Object Detection for Smart Assistive Systems," Journal of AI and Human-Computer Interaction, vol. 14, no. 5, pp. 56-64, 2021.