An Real Time Object Detection Method for Visually Impaired Using Machine Learning

Saravanan Alagarsamy
Department of Computer Science and Engineering.
Kalasalingam Academy of Research and Education,
Anand Nagar, Krishnankoil,
Tamilnadu, India
senthilsar@gmail.com

K. P. L. Syamala
Department of Computer Science Engineering.
Kalasalingam Academy of Research and Education,
Anand Nagar, Krishnankoil,
Tamilnadu, India
9919004138@klu.ac.in

Ch. Sandya Niharika
Department of Computer Science Engineering.
Kalasalingam Academy of Research and Education,
Anand Nagar, Krishnankoil,
Tamilnadu, India
9919004054@klu.ac.in

D. Usha rani
Department of Computer Science Engineering.
Kalasalingam Academy of Research and Education,
Anand Nagar, Krishnankoil,
Tamilnadu, India
9919004068@klu.ac.in

K. Balaji
Department of Computer Science Engineering.
Kalasalingam Academy of Research and Education,
Anand Nagar, Krishnankoil,
Tamilnadu, India
9919004143@klu.ac.in

***Abstract*— Vision, one of the five fundamental human senses, is crucial for defining how people perceive the objects around them. Visual impairments affect more than 200 million people worldwide, severely limiting their ability to perform numerous activities of daily living. Thus, it is essential for blind people to understand their surroundings and the objects they are interacting with. In this work, we created a tool that helps blind persons recognize diverse items in their environment by utilizing the YOLO V3 algorithm combined with R-CNN. This comprises a variety of approaches to develop an app that not only instantly recognizes different objects in the visually impaired person's environment but also guides them using audio output. A convolutional neural network (CNN) called YOLO (You Only Look at Once) recognizes objects in real time. This suggested method is more effective and accurate than other algorithms for recognizing things, according to research results, and it produces results for object detection that are extremely similar in real time. It is crucial for persons who are blind or visually impaired to be able to reliably and effectively detect and recognize objects in order to navigate both common and unfamiliar situations safely, become stronger, and become more independent.**

Keywords*— Convolutional Neural Network, Computer vision, Hyper Text Markup Language

INTRODUCTION

"The eyes are the portals to the soul," It describes the strong connection created when staring someone in the eyes and is crucial to how we view the outside world. The ability to see clearly is crucial for our safety, awareness of our surroundings, and mental alertness [1].

Visually impaired people are blind to the threats they face on a daily basis. They might run across a lot of obstacles while going about their daily lives, even in their cosy surroundings. Humans require vision as a sense since it is essential to how they interpret their surroundings [2].

In Worldwide, 285 million people are believed to have visual impairments, with 39 million being blind and 246 million having impaired vision, according to the World Health Organization (WHO). The number of people with visual impairments is growing as a result of an increase in birth rate, eye conditions, accidents, aging, and other factors [4]. This number rises by up to 2 million people worldwide each year. The ability of the visually impaired to do daily tasks is constrained or negatively impacted. Many people with visual impairments will bring a seeing friend or family member along in order to explore new places. Blind people struggle to interact with others due to these social barriers [5].

Prior research has suggested a number of ways to assist visually impaired people (VIPs) in overcoming their challenges and leading regular lives. These strategies have not adequately addressed the safety concerns for VIPs walking alone, and the offered solutions are frequently complicated, pricey, and unsuccessful, among other things [6].

We suggest an approach based on recent developments in computer vision and machine learning. The YOLO (You Only Look Once) deep learning technology is used to find objects in the immediate vicinity, identify the items, and provide voice output. The camera will take a picture of whatever is in front of the individual [7]. After being processed using deep learning algorithms, the output, which is product identification, is then converted into voice. A method is presented to assist people with vision problems in managing everyday activities like walking, working, and house cleaning. General-purpose object detection should have the following qualities: speed, accuracy, and versatility. The advancement of neural networks has increased the detection frameworks' speed and accuracy. The bulk of detection methods, however, are still restricted to a small number of object classes.

The eyesight of someone who has a visual impairment, also known as impaired vision, is often so badly affected that it cannot be repaired to normal. This implies that a complete rectification is not possible, not even with the aid of glasses, eyeglasses, medication, or eye surgery. Vision impairment might imply different things to different people. The phrase may be used slightly differently by different medical groups, organizations, and practitioners. Even visually impaired people themselves may hold varying views on the subject.

Low vision, which is classified based on the degree of vision loss, is sometimes mistaken for eye issues. The two primary aspects of vision, image quality, and visual field, are usually used to characterize poor vision.

There are numerous potential causes of vision impairment. While some scenarios can take some time to play out and get worse, others might happen quickly. Some people who have bad vision either develop it as they age or are born with it.

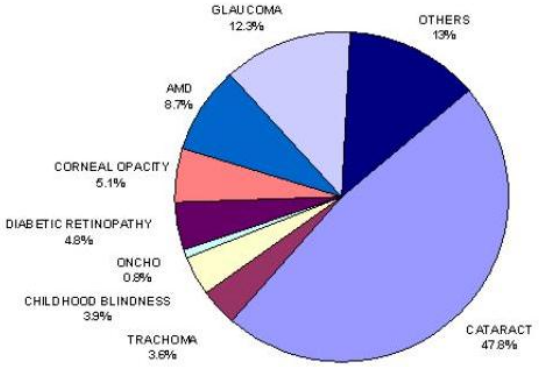

Fig 1: Causes of Visual Impairment

Eye disease - Various levels of irreversible vision loss may be caused by particular eye conditions. Illness - Diseases in a number of different physiological systems can also impair vision. These problems can harm your vision in a variety of ways. Injury - Head injuries have the potential to permanently or completely damage eyesight.

RELATED WORKS

S.No	Author details	Used methodology	Advantages and limitations
1	Redmon et al. [3]	Multi-box detector algorithm	The authors discussed the accuracy validation parameters aspect ratio as well as how they increased the accuracy in their convolutional layers using depth- and space-wise separable convolutional layers. The single shot multi-box detector (SSD) algorithm, which uses a single layer of a neural network to detect objects in the image, was also suggested by the authors as the fastest optimization technique.
2	Girshick et al [8]	Microsoft COCO: Common Objects in Context.	This work suggested a new dataset with the intention of improving the state-of-the-art in object recognition by placing the problem of object recognition in the context of the more general problem of pattern recognition. With 328k photos in the dataset containing photographs of common things, there are a total of 2.5 million instances in which each object has been labeled, making the object detection problem simpler to solve.
3	Kumar et al. [9]	You Only Look Once: Unified, Real-Time Object Detection	This research shows how Fast YOLO can process an astounding 155 frames per second while still obtaining twice the mAP of other real-time detectors. On the baseline, modern detection techniques produce fewer false positive predictions, while YOLO commits more localization mistakes. Last but not least, YOLO chooses very general examples. It performs better than other detection methods like DPM and R-CNN when generalizing from natural images when used in other domains, such as art.
4	Ramík et al. [10]	Detection Based on CNN	According to this study, YOLOv2 performs better than Faster R-CNN in terms of performance and accuracy and also features an object detector with excellent generalization properties that can represent the full image.
5	Ren et al. [11]	Blind Person Assistant: Object Detection	The TensorFlow Object Detection API allows for the detection of several objects. They presented an algorithm in this publication (SSD). SSD uses a similar phase during training to connect the proper anchor box with the bounding boxes of each underlying data object within an image.
6	Saini et al. [12]	SSD for Real-Time Pill Identification	This study finds that YOLO v3 can be used for real-time object recognition because it has faster detection speeds than R-CNN and SSD while maintaining a specific MAP.
7	Arafat et al. [13]	Object Detection System for Visually Impaired Persons Using Smartphones	Building a way to recognize objects in real-time utilizing a camera as an input device and communicating that information to the user via a smartphone and headphones is the aim of this project. The system would use an audio device, such as speakers or headphones, to deliver information about products in order to aid persons who are visually impaired.

After performing the survey, the combination of R-CNN with Yolo is identified to improve the performance of the algorithm in identifying the objects exactly for blind people with the help of an android mobile phone.

PROPOSED SYSTEM

The development of portable assistive technology systems aims to improve the capacities of people with disabilities. One of a person's most important senses is vision. Vision is the most important sense that helps us be aware of how we interpret our surroundings [14]. Legally blind people frequently struggle to understand what is going on around them, especially in an outdoor setting where things are continuously moving and shifting. Approaches for object classification would go a long way toward assisting visually impaired individuals in navigating the challenges they face every day. The object-detecting system's objective is to provide visually impaired individuals with a simple, approachable, practical, inexpensive, and efficient method. With the help of a smartphone, headphones, and an input device like a camera, this system aims to develop an application that can recognize objects in real-time. The application would use an audio recorder, such as speakers, to convey information about products as voice output and increase the independence of persons who are visually impaired. The suggested method helps people who are visually impaired identify and steer clear of objects in both indoor and outdoor settings that obstruct everyday duties and job motivation. People with vision impairment would find daily life much easier if they were aware of the stuff nearby [15].

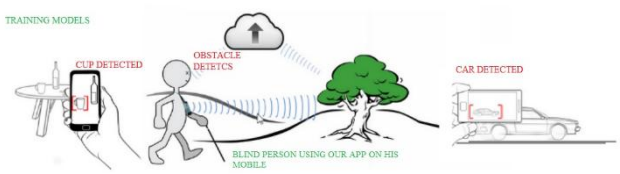

Fig.2. Blind Person Using App on his Mobile

In this project, an Android application that can recognise objects will be developed for users with vision impairments. Due to the specially designed application, it is intended for blind people to find the goods they are looking for in daily life more quickly. The application will benefit from object detection technology that is based on vision.

For the advantage of the blind individual, an effort will be made to gauge the object's location and distance during object detection. A voice message including the names and locations of the recognized objects will then be delivered to the visually impaired person.

Additionally, the user will receive a voice message describing the functioning of the application, making it simple for blind individuals to utilize. The software will then be tested to determine how well it can recognize objects and behave in various situations.

The brain's neural networks are artificially created. They function similarly to the brain. Convolutional neural networks are the most significant neural network utilized for image recognition. Multiple layers of neurons in convolutional neural networks are used to extract information from the image. The model grows complex without convolution layers since they need a lot of neurons depending on the size of the image. Convolution layers apply convolution operations to every layer in order to create a feature map and transmit the results to the following layer. In these levels, feature extraction is carried out [16].

The forward propagation and backward propagation, which are combined to form an epoch, are the two crucial events that take place throughout the model's training. An era is a single cycle of forward and backward propagation. From the input layer to the output layer, forward propagation travels in a single direction. The feature extraction, weight computation, and application of the activation function take place in forward propagation, after which the error is computed. The weights will be adjusted in the backward propagation method, which moves from the input layer to the output layer in a reverse manner, based on the error value. The number of epochs is represented by the number n at which this process occurs [17].

The biggest issue that arises when these methods are applied in real time is the FPS value. The frame rate of an object detection model determines how quickly the video is processed and output is produced. FPS is a metric used to describe how quick a procedure is. For the real-time effect, the FPS value must be higher than 20. However, the CNN family algorithms' highest FPS value is only 18 (faster R-CNN). The mAP value is a typical performance statistic (mean average precision value).

As there will only be one forward propagation to identify objects in each run, the YOLO (You Only Look Once) method enhanced FPS and mAP values while reducing run time complexity. The FPS value margin of 155 was attained by this algorithm. It also comprehends how items are represented generally. Each version's backbone was a neural network, such as DarkNet or efficientNet. It gained notoriety for its accuracy, quickness, and capacity for learning.

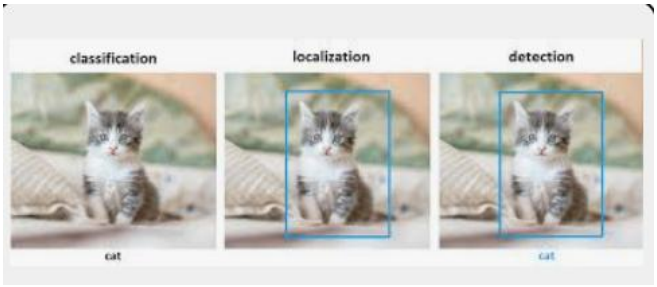

Fig.3.Detection process of Yolo

Fig.3 epitomizes the detection process of YOLO. Grids split each cell, and it is the duty of each cell to locate the object. Three techniques—Residual blocks, Bounding boxes, and Intersection over Union—are the foundation of the algorithm. IoU (intersection over union) is another evaluation statistic that shows how much two boxes overlap. It assessed how well original and anticipated values coincided. It accepts input in the form of weights and coordinates for training, unlike many neural algorithms.

(A) YOLO Architecture

YOLO is a fast-working method with good accuracy and FPS values that can be utilized for real-time object detection in comparison to many other existing techniques. The object

detection model, which was based on the DarkNet53 model, was constructed using the YOLOv3 version.

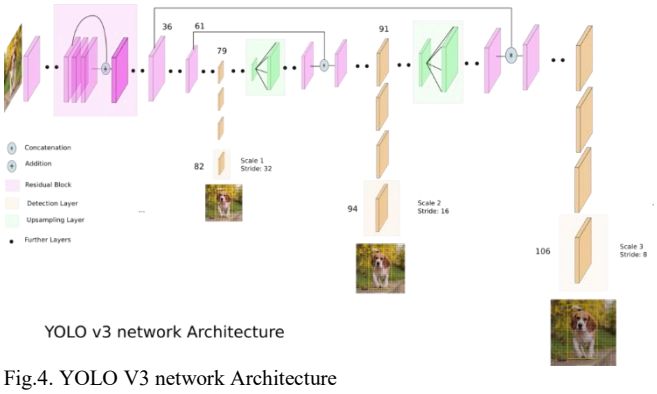

Fig.4. YOLO V3 network Architecture

Fig.4 epitomizes the architecture of YOLO V3. The utilization of the method is combined with the R-CNN networks. YOLO model is built, and Weights are assigned based on the feature of the images. The shapes of the images are predicted using bound box method. Edges of object is used for identifying the different entity.

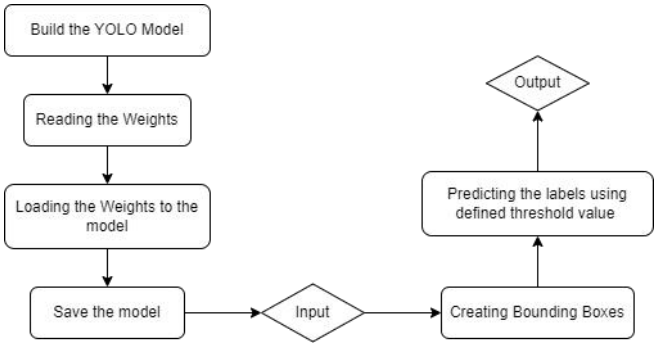

Fig.5.Prediction of the target variable

Fig. 5 embodies the prediction process of the target variables. The output of the YOLO technique is combined with the R-CNN method to increase the accuracy of the prediction process. Based on the predetermined feature set, the object is predicted.

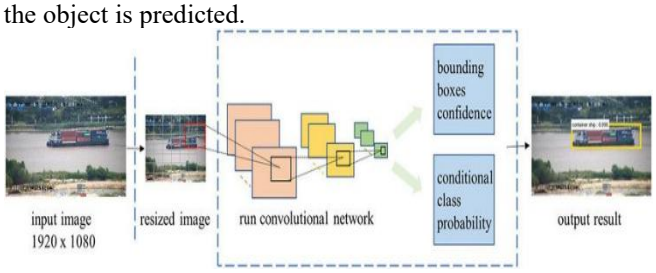

Fig.6.Working principle of R-CNN with YOLO method

Fig.6. depicts the working principles of the suggested method. After the model is detected the image, then the image is fed into the CNN model. The model prediction is performed using the CNN model.

(B) R-CNN combined with Yolo

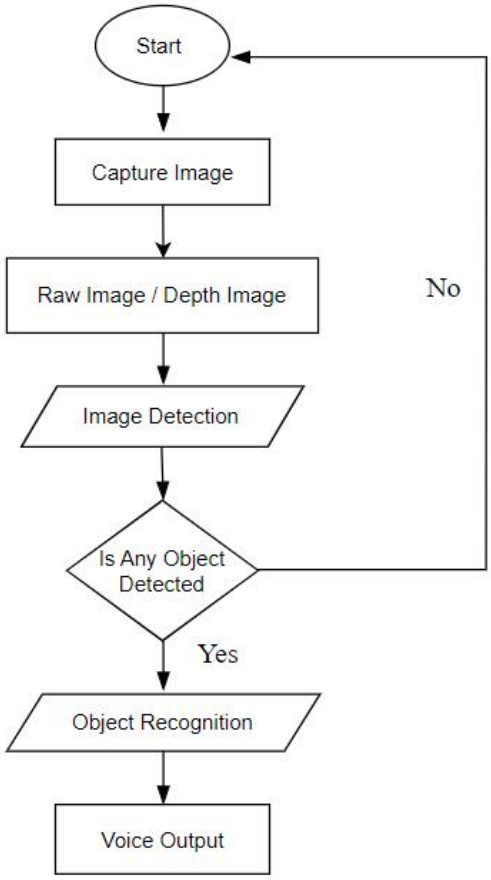

Fig.7.Flowchart of the proposed system

Fig.7 exemplifies the complete process of the suggested method. The application must have access to the mobile device's camera, speaker, and microphone in order to work as intended, as indicated in the diagram. The user starts the application first. The device's camera and microphone are then requested from the user. If accepted, the procedure is carried out; otherwise, the application is rejected. The novelty of the suggested method is emphasized based on producing the audio output of the predicted object. The object detected by the model will be delivered to the user in the format of voice. This processing aid for blind people to assist and travel to many places.

RESULTS AND DISCUSSION

The data set used for testing is VOC 2012 real-time test data also the error analysis of the suggested method is performed based on the above data set. First, the model is tested with the wildlife animals. The animals are detected based on the predefined shapes fed into the model.

Fig.8 indicates the prediction of wildlife animals. Edge detection is used to capture the structure of the image. The boundary of the image is detected then the object is mapped with the features data set fed.

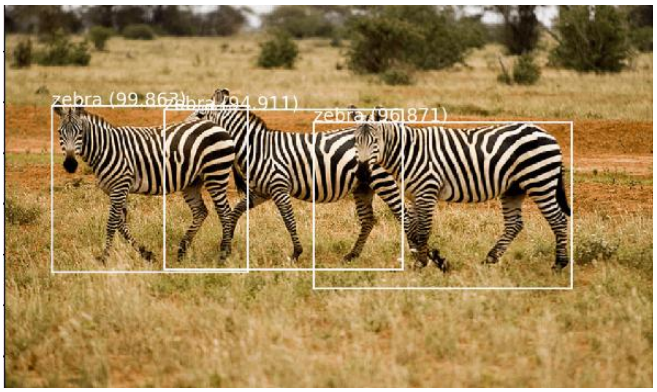

Fig.8.Wildlife animal prediction

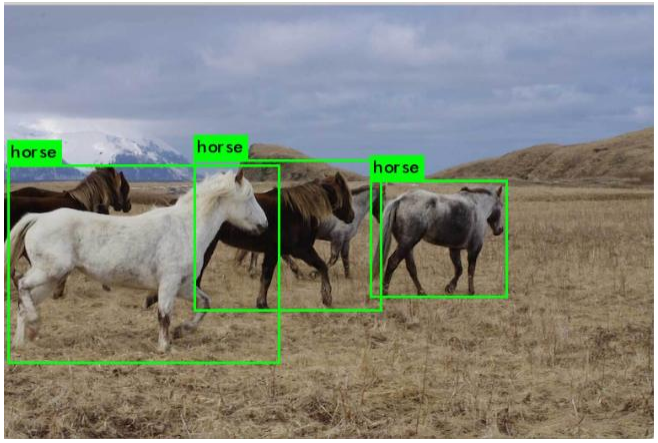

Fig.9.Object detection of R-CNN with YOLO

Fig.9. embodies the process of object detection of the suggested method. The detected image is mapped with the predefined image. If a match is found, the output will be delivered to the user in the format of voice for recognition.

TABLE1. ACCURACY COMPARISON

S. No	State of art methods	MAP in %
1	R-CNN+Yolo	96.48
2	CNN	90.23
3	Yolo	89.36
4	RCNN	91.26
5	Hypernet	84.45
6	Edge based detection	88.78

Table 1 encapsulates the accuracy of the suggested and comparative techniques. The feature set map of the suggested and related techniques is presented in fig.9.Based on the feature set mapped to the model, the accuracy of the method will deliver. The R-CNN +Yolo deliberately improved accuracy in finding the objects based on the predefined shapes. The sample type of object is trained with different shapes in various dimensions.

Fig.9 reveals the accuracy comparison. The suggested model provides more accurate information while comparing it with the related methods.

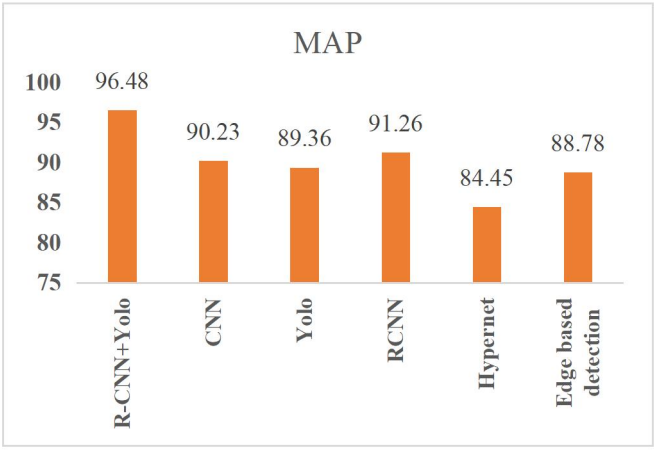

Fig.9 MAP comparison

Fig.9 reveals the accuracy comparison. The suggested model provides more accurate information while comparing it with the related methods.

V. CONCLUSION

The suggested method act as a tool that aids blind people in recognizing various objects in their environment. This includes a number of techniques to create an app that not only instantaneously recognizes various items in the environment of the visually impaired person, but also directs them through auditory output. Real-time object recognition is accomplished with an RCNN with YOLO. The suggested method is more efficient and accurate than other algorithms for object recognition, and it delivers results for object detection that are quite similar in real-time.

REFERENCES

1. J. -G. Kim and J. -H. Yoo, "HW Implementation of Real-Time Road & Lane Detection in FPGA-Based Stereo Camera," 2019 IEEE International Conference on Big Data and Smart Computing (BigComp), 2019, pp. 1-4.
2. J. Hosang, R. Benenson, P. Dollár and B. Schiele, "What Makes for Effective Detection Proposals?," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 4, pp. 814-830, 1 April 2016.
3. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
4. K. G, R. P, N. N N, A. Beulah and R. Priyadharshini, "Detection of Electronic Devices in real images using Deep Learning Techniques," 2021 5th International Conference on Computer, Communication and Signal Processing (ICCCSP), 2021, pp. 295-300.
5. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition" in Computer Vision and Pattern Recognition, IEEE, pp. 770-778, 2016..
6. P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, Sept. 2010.
7. P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627-1645, Sept. 2010.
8. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 580-587.

9. R. Kumar, S. Lal, S. Kumar and P. Chand, "Object detection and recognition for a pick and place Robot," Asia-Pacific World Congress on Computer Science and Engineering, 2014, pp. 1-7.
10. Ramík, D.M., Sabourin, C., Moreno, R. and Madani, K., 2014. A machine learning based intelligent vision system for autonomous object detection and recognition. Applied intelligence, 40(2), pp.358-375.
11. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017.
12. S. S. Saini and P. Rawat, "Deep Residual Network for Image Recognition," 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2022, pp. 1-4.
13. S. Y. Arafat and M. J. Iqbal, "Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning," in IEEE Access, vol. 8, pp. 96787-96803, 2020.
14. W. Liu, G. Wu, F. Ren and X. Kang, "DFF-ResNet: An insect pest recognition model based on residual networks," in Big Data Mining and Analytics, vol. 3, no. 4, pp. 300-310, Dec. 2020.
15. X. Li, L. Ding, L. Wang and F. Cao, "FPGA accelerates deep residual learning for image recognition," 2017 IEEE 2nd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017, pp. 837-840.
16. Y. Li and H. Chen, "Image recognition based on deep residual shrinkage Network," 2021 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), 2021, pp. 334-337.
17. Z. Zhang, S. Qiao, C. Xie, W. Shen, B. Wang and A. L. Yuille, "Single-Shot Object Detection with Enriched Semantics," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5813-5821.