Smart Vision for the Blind- YOLOv8-Driven Object Detection and Recognition Framework

Prof. D. V. Varaprasad, M.Tech, (Ph.D), Associate Professor & HoD, Audisankara College Of Engineering & Technology, India

V.Sreenatha Sarma, Professor, Dean ICT,Department of CSE, Audisankara College Of Engineering & Technology, India

Gurram V N Koteswara Rao, Department of CSE, Audisankara College Of Engineering & Technology, India

ABSTRACT_ Blindness is one of the top ten disabilities in humans, and India unfortunately has the world's largest visually impaired population. In this paper, we present a novel framework for assisting the visually impaired in object detection and recognition, allowing them to navigate and be aware of their surroundings independently. The paper employs a Yolo v8mechanism for object detection and classification, followed by human face and currency note recognition.Yolo v8 detector is trained on the coco dataset, and a new class is added to detect currency as well. Finally, the framework's output can be presented to the visually impaired person in audio format. The mean Accuracy and Precision (mAP) scores of the standalone Yolo v8detector of the added currency class were 67.8 percent, and the testing accuracy of person and currency recognition was 90 to 92 percent.

1.INTRODUCTION

Outwardly disabled individuals face a ton of challenges in their lives. According to the World Health Organization (WHO) in 2019, approximately 2.2 billion people worldwide suffer from vision impairment. For people with visual impairments, recognizing and spotting common objects in their surroundings appears to be a monumental task. They either rely on other people, making the blind dependent on them, or on their sense of touch and smell, which is very inaccurate and can be dangerous in some situations, to detect objects. The most widely used blind navigational aid is the white cane. This was additionally improved by adding ultrasonic and IR sensors to recognize snags nearby the outwardly impeded client, and give criticism as vibration or sound. Even though this method helped the visually impaired person move around, it didn't tell them much about the surrounding area. Objection detection and classification, followed by recognition and audio feedback, are crucial for the user's comprehension of the environment.

Convolutional neural networks, in particular, have demonstrated promising results in image-based object detection, classification, and recognition tasks. The authors of [1] use a feed-forward neural network to suggest products for shopping in speech. In [2], an obstacle detection and classification system based on smartphones is implemented. The multiscale Lucas-Kanade algorithm is used to extract and track interest points; homographic transforms and agglomerative clustering are used to estimate background motion; and the Histogram of Oriented Gradients (HOG) descriptor is used to classify the data into a Bag of Visual Words (BoVW). [3] presents a survey on Electronic Travel Aids (ETA), which are navigational aids for the visually impaired. The advantages and disadvantages of various ETAs are discussed and compared feature-by-feature. It additionally features the way that no ongoing framework consolidates every vital component and any innovation shouldn't endeavor to supplant the stick however to supplement it by legitimate alarming and criticism. The most recent methods for activity recognition are outperformed by a deep novel architecture for the visually impaired that makes use of a late fusion of two parallel CNNs [4]. The two CNN's GoogLeNet and AlexNet complete one another in recognizing various elements of a similar class, subsequently the info video is taken care of to the two of them, and the result class scores are joined utilizing Backing Vector Machine (SVM). For object detection and color recognition, another novel approach proposed in [5] employs CNN, a recurrent neural network (RNN), a softmax classifier, and hue, saturation, and intensity (HSI) color thresholding. [6] demonstrates a strategy for a visually impaired outdoor navigation assistant that combines computer vision and deep learning methods. The system handles sudden camera movements, employs You Only Look Once (YOLO) for object recognition, and uses a regression-based mechanism for object tracking without prior information. In [7], a smartphone app is made to help visually impaired people navigate. It has two modes of operation: online and offline based on how connected the user's network is. The web-based mode involves Quicker RCNN to produce expectations in stable circumstances and Consequences be damned for quicker results. In contrast, in offline mode, a feature recognition module employing Haar features and a Histogram of Gradients (HOG) serves this purpose. A CNN is intended for pre-prepared object acknowledgment utilizing the ImageNet dataset [8]. An original DLSNF (Profound Learning-based Tangible Route Structure) based on the Just go for it engineering is proposed in [9] for planning a tactile route gadget on top of NVIDIA Jetson TX2. SqeezeNet, a light-weight pre prepared CNN model, accomplished better execution and decreased computational dormancy per picture [10]. The last convolutional layer's weights were changed, the activation function was changed from Rectified Linear Unit (ReLU) to LeakyReLU, and a batch normalization layer was added to SqueezeNet.

2.LITERATURE SURVEY

[1] Ivo M. Baltruschat, Hannes Nickisch, Michael Grass, Tobias Knopp, and Axel Saalbach, "Comparison of deep learning approaches for multi-label chest X-ray classification," Scientific Reports, vol. 9, no. 1, Apr 2019.

The ChestX-ray14 dataset and other labeled X-ray image archives have made deep learning methods more widely accessible. To give better knowledge into the various methodologies, and their applications to chest X-beam characterization, we examine a strong organization engineering exhaustively: a ResNet-50 We consider transfer learning with and without fine-tuning, as well as the training of a dedicated X-ray network from scratch, building on previous research in this field. To use the high spatial goal of X-beam information, we likewise incorporate a lengthy ResNet-50 engineering, and an organization coordinating non-picture information (patient age, orientation and securing type) in the grouping system. In a finishing up try, we likewise explore different ResNet profundities (for example ResNet-38 and ResNet-101). Using 5-fold resampling and a multi-label loss function, we conduct a systematic evaluation of the various pathology classification ROC statistics methods and rank correlation differences between classifiers. Generally, we notice a significant spread in the accomplished presentation and presume that the X-beam explicit ResNet-38, coordinating non-picture information yields the best by and large outcomes. Class activation maps are also used to comprehend the classification procedure, and the impact of non-image features is thoroughly examined.

[2] Z. Xue, S. Candemir, S. Antani, L. R. Long, S. Jaeger, D. Demner-Fushman, and G. R. Thoma, "Foreign object detection in chest X-rays," in 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Nov 2015, pp. 956–961.

One important method for screening and identifying pulmonary diseases is automatic analysis of chest X-ray images.

The presence of unfamiliar items in the pictures obstructs the exhibition of such handling. In this paper, we concentrate on a specific kind of foreign object that frequently appears in the images of a large dataset of chest X-rays: the buttons on the patient's gown. There are four major steps in our proposed method: segmentation of lung regions, button object extraction, low contrast image identification and enhancement, and intensity normalization. We used two approaches for the step of button object extraction based on the characteristics of the button objects. One depended on the round Hough change; The Viola-Jones algorithm was the foundation of the other. We tried and looked at the two strategies utilizing a ground truth dataset containing 505 button objects. The outcomes demonstrate the proposed method's efficacy.

[3] James P. Howard, Louis Fisher, Matthew J. Shun-Shin, Daniel Keene, et al., "Cardiac rhythm device identification using neural networks," JACC: Clinical Electrophysiology, vol. 5, no. 5, pp. 576 – 586, 2019.

This paper reports the turn of events, approval, and public accessibility of another brain network-based framework which endeavors to distinguish the producer and, surprisingly, the model

4.RESULTS AND DISCUSSION

gathering of a pacemaker or defibrillator from a chest radiograph. Clinical staff frequently need to decide the model of a pacemaker or defibrillator (heart musicality gadget) rapidly and precisely. Current methodologies include contrasting a gadget's radiographic appearance and a manual stream diagram.

3.PROPOSED SYSTEM

Object detection is the principle objective of this system. It consists of object classification and object localization.

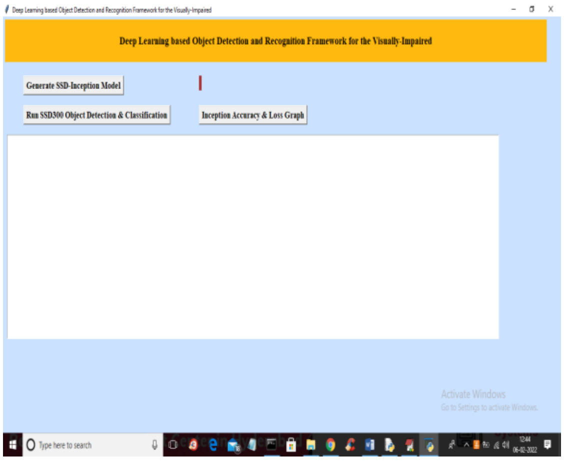
- The application is build to recognize or detect some objects like chair, mobile etc and some outdoor objects like vehicles, people, etc.
- The application will use laptop camera to scan the surrounding in real time and take the frames from the ongoing video..

3.1 IMPLEMENTATION

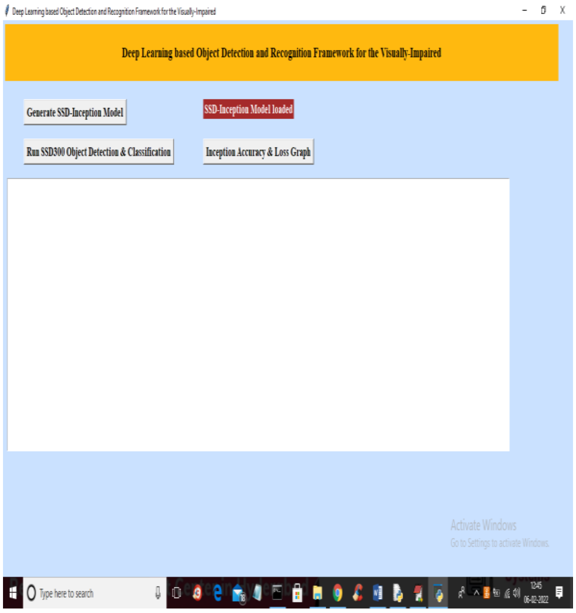
1) Generate SSD-Inception Model: using this module we will generate and load SSD300 and inception model

2) Run SSD300 Object Detection & Classification: using this model we will detect and classify object (21 classes with currency) using SSD and Inception model

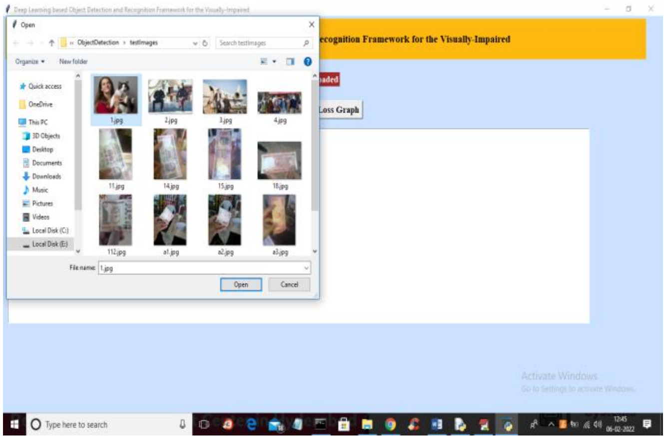
3) Inception Accuracy & Loss Graph: using this module we will plot Inception training accuracy and loss graph



In above screen click on 'Generate SSD-Inception Model' button to load models and to get below screen

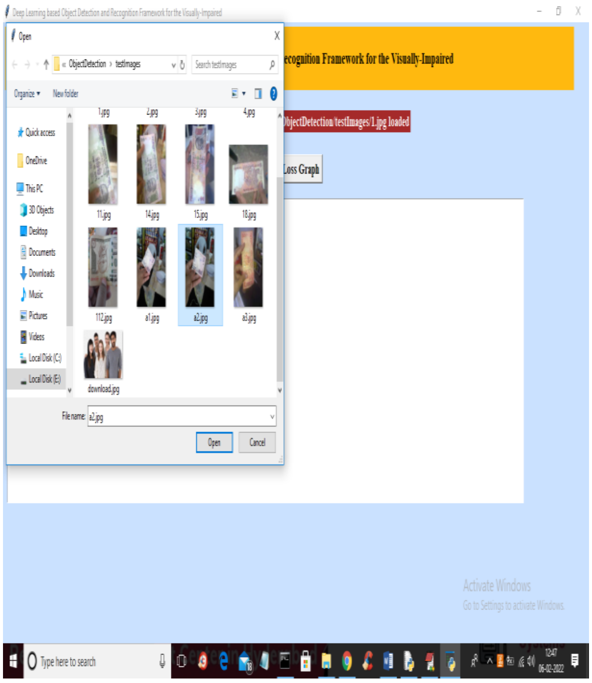


In above screen in red colour text we can see models loaded and now click on 'Run SSD300 Object Detection & Classification' button to upload image and then classify object



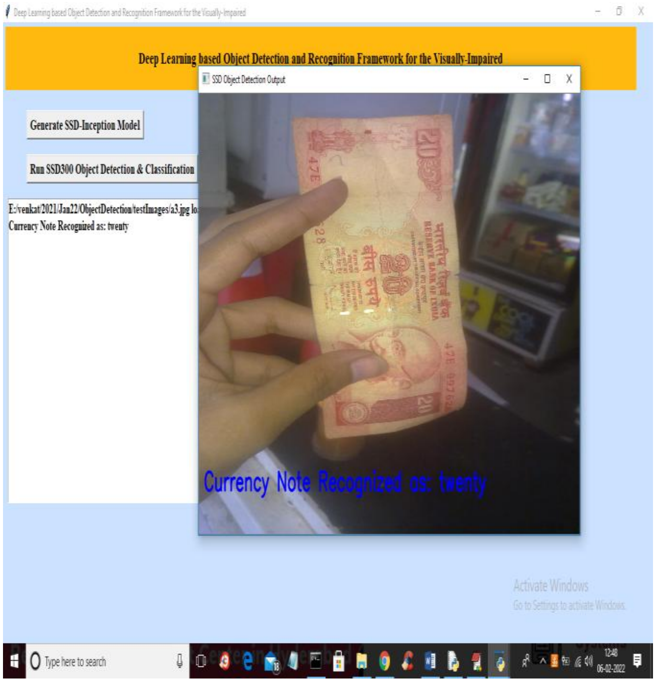
In above screen selecting and uploading 1.jpg file and then click on 'Open' button to get below output

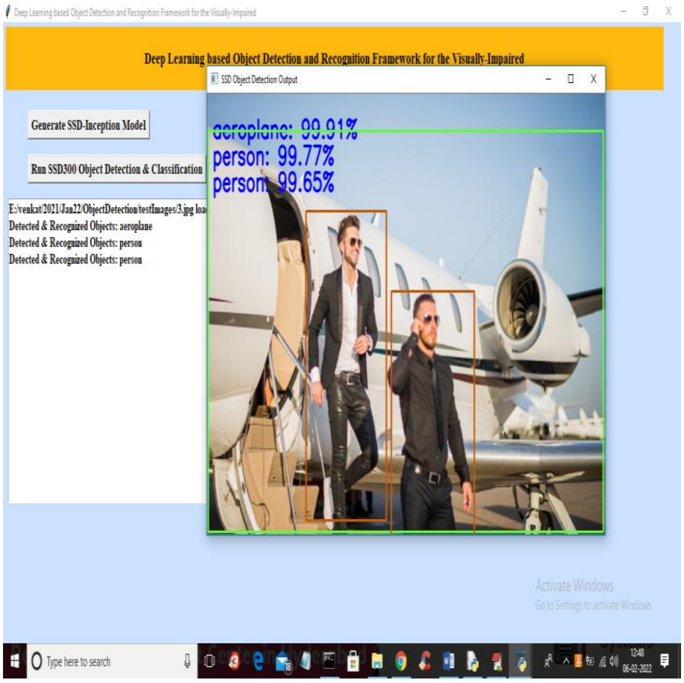
In above screen objects detected and classified as person and cat and now try other images



For above a2.jpg selection below is the output

In above screen application displaying chair and currency note fifty detected and no test other images





In above screen in text area as well as images you can see printing of detected objects and similarly you can upload and test other images

5.CONCLUSION

To assist visually impaired people, a novel framework employing object detection, classification, and face and currency recognition has been presented. Once the training is completed, it is fairly simple and easy to deploy. Using separate Inception models for face and currency recognition speeds up the process and makes it more user-specific and adaptable. It is one of the most generic frameworks, incorporating all useful features, and will undoubtedly be of great service to humanity. More work can be done in the future to make the face and currency recognition spoof-proof..

REFERENCES

[1] F. Jabeen, A. Muhammad, and A. M. Enriquez, "Feed forward neural network training based interactive shopping for blind," in 2015 12th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE). IEEE, 2015, pp. 1–6.

[2] R. Tapu, B. Mocanu, A. Bursuc, and T. Zaharia, "A smartphonebased obstacle detection and classification system for assisting visually impaired people," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013, pp. 444–451.

[3] R. Tapu, B. Mocanu, and E. Tapu, "A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments," in 2014 11th international symposium on electronics and telecommunications (ISETC). IEEE, 2014, pp. 1–4.

[4] J. Monteiro, J. P. Aires, R. Granada, R. C. Barros, and F. Meneguzzi, "Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks," in 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, 2017, pp. 2267–2274.

[5] R. Kumar and S. Meher, "A novel method for visually impaired using object recognition," in 2015 International Conference on Communications and Signal Processing (ICCSP). IEEE, 2015, pp. 0772–0776.

[6] R. Tapu, B. Mocanu, and T. Zaharia, "Seeing without sight-an automatic cognition system dedicated to blind and visually impaired people," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017, pp. 1452–1459.

[7] B.-S. Lin, C.-C. Lee, and P.-Y. Chiang, "Simple smartphone-based guiding system for visually impaired people," Sensors, vol. 17, no. 6, p. 1371, 2017.

[8] K. Potdar, C. D. Pai, and S. Akolkar, "A convolutional neural network based live object recognition system as blind aid," arXiv preprint arXiv:1811.10399, 2018.

[9] J.-C. Ying, C.-Y. Li, G.-W. Wu, J.-X. Li, W.-J. Chen, and D.-L. Yang, "A deep learning approach to sensory navigation device for blind guidance," in 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018, pp. 1195–1200.