DEEP LEARNING BASED OBJECT DETECTION AND RECOGNITION FRAMEWORK FOR THE VISUALLY-IMPAIRED

[1]P. HYMAVATHI, [2] K. ARCHANA, [3] GOWTHAMI

[1,2, 3] Assistant Professor, Department of Computer Science and Engineering, Sri Indu College of Engineering and Technology, Hyderabad, Telangana-501510

ABSTRACT

Vision impairment or blindness is one of the top ten disabilities in humans, and unfortunately, India is home to the world's largest visually impaired population. In this study, we present a novel framework to assist the visually impaired in object detection and recognition, so that they can independently navigate, and be aware of their surroundings. The paper employs transfer learning on Single-Shot Detection (SSD) mechanism for object detection and classification, followed by recognition of human faces and currency notes, if detected, using Inception v3 model. SSD detector is trained on modified PASCAL VOC 2007 dataset, in which a new class is added, to enable the detection of currency as well. Furthermore, separate Inception v3 models are trained to recognize human faces and currency notes, thus making the framework scalable and adaptable according to the user preferences. Ultimately, the output from the framework can then be presented to the visually impaired person in audio format. Mean Accuracy and Precision (mAP) scores of standalone SSD detector of the added currency class was 67.8 percent, and testing accuracy of person and currency recognition of Inception v3 model were 92.5 and 90.2 percent respectively.

INTRODUCTION:

Visually impaired people face a lot of difficulties in their lives. Recent statistics published by World Health Organization (WHO) in 2019 reveal that globally, around 2.2 billion individuals are affected by vision impairment. Detecting and recognizing common objects in the surroundings seem to be a herculean task for the visually impaired individuals. They rely either on other people, which makes the blind dependent on them, or, on their sense of touch and smell to detect objects, which is highly inaccurate and can be hazardous in some cases [1].

The white cane is the most popular blind navigating device. This was further improved by adding ultrasonic and IR sensors to detect obstacles in the vicinity of the visually impaired user, and provide feedback in the form of vibration or sound. Though this approach was useful for the mobility of the visually impaired user, it provided little or no information about the surroundings. For the user to have a better understanding of the surrounding, objection detection and classification, followed by recognition and audio feedback is crucial [2].

Neural networks, particularly, convolutional neural networks have shown promising results particularly in object detection, classification, and recognition tasks from images. The authors use a feed-forward neural network to provide speech suggestions regarding products of shopping. Real Time smartphone-based obstacle detection and classification system is implemented. The detection process involves interest point extraction and tracking through

multiscale Lucas - Kanade algorithm, background motion estimation using homographic transforms and agglomerative clustering technique, followed by classification with the help of Histogram of Oriented Gradients (HOG) descriptor into Bag of Visual Words (BoVW). A survey on Electronic Travel Aids (ETA) designed for visually impaired navigation assistance is presented. Various ETAs, their strengths, and shortcomings are discussed and compared feature-wise. It also highlights the fact that no current system incorporates all necessary features and any technology should not attempt to replace the cane stick but to complement it by proper alerting and feedback [3].

A deep novel architecture for visually impaired employing a late fusion of two parallel CNN's outperforms the state-of the-art methods for activity recognition [18]. The two CNN's GoogLeNet and AlexNet complement each other in identifying different features of the same class, hence the input video is fed to both of them, and the output class scores are combined using Support Vector Machine (SVM). Yet another novel method proposed uses CNN followed by a recurrent neural network (RNN) and softmax classifier for object detection, and Hue, Saturation and Intensity (HSI) color thresholding for color recognition. An approach combining computer vision and deep learning techniques for visually impaired outdoor navigation assistant is shown. The system uses a regression-based mechanism for object tracking without a priori information, handles sudden camera movements, and exploits You Only Look Once (YOLO) for object recognition [4].

A smartphone app is designed for guiding visually impaired persons. It can operate in two modes: online and offline based on user network connectivity [16][17]. The online mode uses Faster RCNN to generate predictions in stable conditions and YOLO for faster results. Whereas, a feature recognition module using all the features and Histogram of Gradients (HOG) serves this purpose in offline mode. A CNN is designed for pre-trained object recognition using the ImageNet dataset. A novel DLSNF (Deep-Learning-based Sensory Navigation Framework) built on the YOLO architecture is proposed for designing a sensory navigation device on top of NVIDIA Jetson TX2. SqeezeNet, a light-weight pretrained CNN model, achieved better performance and reduced computational latency per image. SqueezeNet is improved by changing the weights of the last convolutional layer, replacing the Rectified Linear Unit (ReLU) with LeakyReLU as activation function and addition of batch normalization layer [5].

LITERATURE SURVEY

Feed forward neural network training based interactive shopping for blind

Today shopping markets pay attention towards customer needs and services. Unfortunately, the blind and vision impaired person are still incapable to access these environments without reliance [14]. Assistive technology is trying to sway the living style of the blind by introducing support systems for routinely actions like reading, writing, walking, Web surfing, and shopping. However, still the blind have to count on others for personal accessories shopping [15]. To overcome this problem, we designed and developed a feed forward talking accessories selector. Our system is trained using feed forward techniques with a feature level block-based multi-focus image fusion method to provide suggestions regarding product selection, fitness, and color combination, for instance in dress and jewelry. The evaluation of our system takes into account specialists opinions, such that statistical analysis shows similarity between both [6].

A smartphone based obstacle detection and classification system for assisting visually impaired people

In this paper we introduce a real-time obstacle detection and classification system designed to assist visually impaired people to navigate safely, in indoor and outdoor environments, by handling a smartphone device. We start by selecting a set of interest points extracted from an image grid and tracked using the multiscale Lucas - Kanade algorithm. Then, we estimate the camera and background motion through a set of homographic transforms. Other types of movements are identified using an agglomerative clustering technique. Obstacles are marked as urgent or normal based on their distance to the subject and the associated motion vector orientation [7]. Following, the detected obstacles are fed/sent to an object classifier. We incorporate HOG descriptor into the Bag of Visual Words (BoVW) retrieval framework and demonstrate how this combination may be used for obstacle classification in video streams. The experimental results demonstrate that our approach is effective in image sequences with significant camera motion and achieves high accuracy rates, while being computational efficient [8].

A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments

A survey on wearaIn this paper we introduce a comprehensive survey of wearable systems designed to assist the visual impaired users navigation in everyday life outdoor scenarios. We focus on presenting the main advantages and limitations of each technique in effort to inform the scientific community about the progress in the area of assistive devices and also offer users a review about the capabilities of each system. Various performance parameters are introduced in order to classify different systems by giving qualitative and quantitative measures for evaluation. At the end of the study conclusions are presented along with some perspectives for future work and development. The devices used to assist the visual impaired user navigation in outdoor environments [9].

SYSTEM ANALYSIS AND DESIGN

Existing System & its Disadvantages:

Vision impairment or blindness is one of the top ten disabilities in humans, and unfortunately, India is home to the world's largest visually impaired population. In this study, we present a novel framework to assist the visually impaired in object detection and recognition, so that they can independently navigate, and be aware of their surroundings. The paper employs transfer learning on Single-Shot Detection (SSD) mechanism for object detection and classification, followed by recognition of human faces and currency notes, if detected, using Inception v3 model. SSD detector is trained on modified PASCAL VOC 2007 dataset, in which a new class is added, to enable the detection of currency as well. Furthermore, separate Inception v3 models are trained to recognize human faces and currency notes, thus making the framework scalable and adaptable according to the user preferences [10].

Disadvantages:

• Increased Dependency: Visually-impaired individuals may become more dependent on others for assistance in identifying objects and navigating their environment of the limiting their independence and autonomy.

• Safety Risks: Without an object detection and recognition system, visually-impaired individuals may be more prone to accidents and injuries due to obstacles and hazards that they are unable to detect [11].

Proposed System & its Advantages:

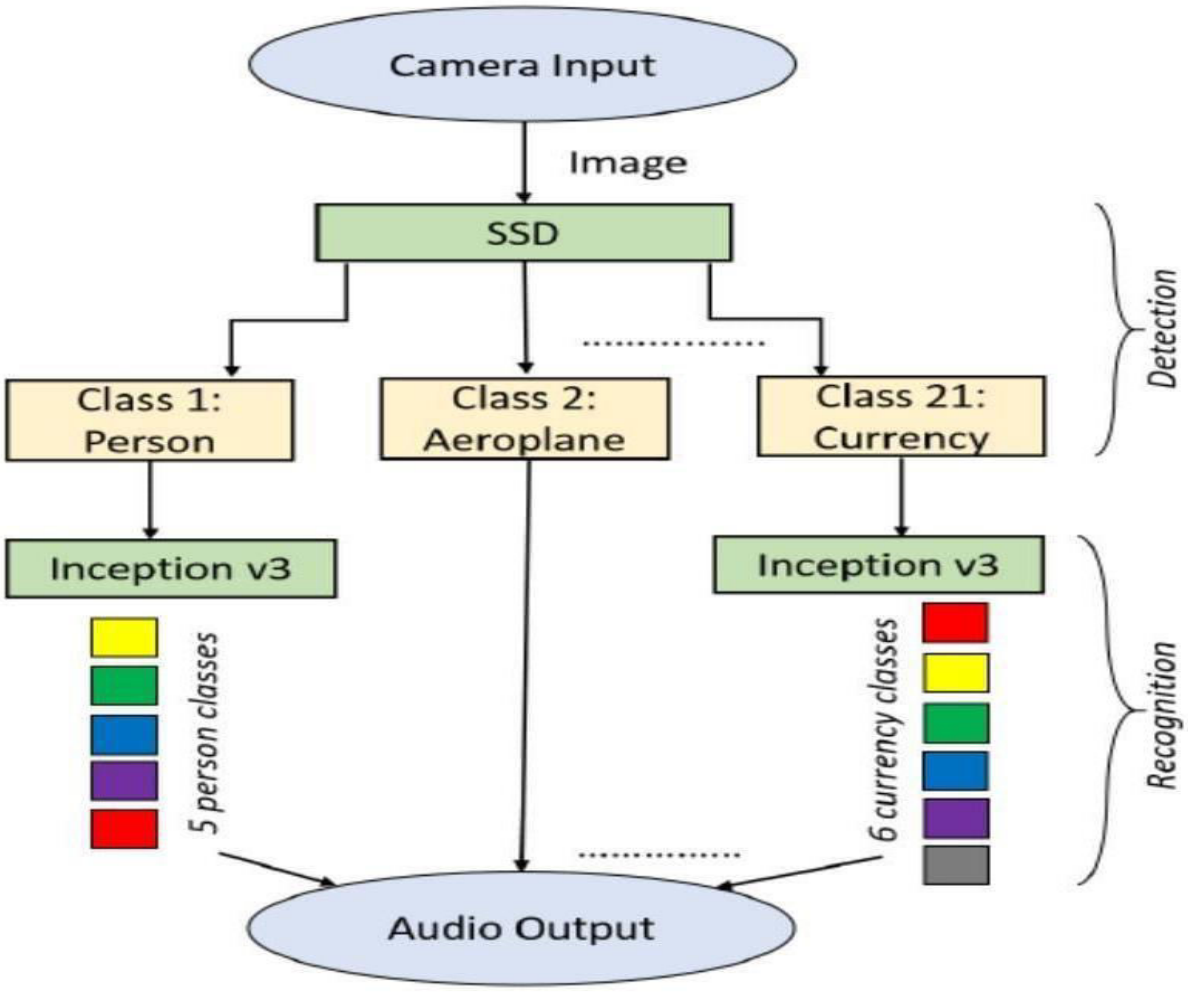
A deep novel architecture for visually impaired employing a late fusion of two parallel CNN's outperforms the state-of the-art methods for activity recognition [4]. The two CNN's GoogLeNet and AlexNet complement each other in identifying different features of the same class, hence the input video is fed to both of them, and the output class scores are combined using Support Vector Machine (SVM). Yet another novel method proposed in [5] uses CNN followed by a recurrent neural network (RNN) and softmax classifier for object detection, and Hue, Saturation and Intensity (HSI) color thresholding for color recognition. An approach combining computer vision and deep learning techniques for visually impaired outdoor navigation assistant [12].

Advantages:

• Increased Independence: With the help of an object detection and recognition system, Visually-impaired individuals can be more independent in navigating their environment. They can identify objects and obstacles without relying on assistance from others [13].

• Improved Safety: By detecting obstacles and hazards, the system can help prevent accidents and injuries. This is especially important for visually - impaired individuals who may be at higher risk of falls and other accidents.

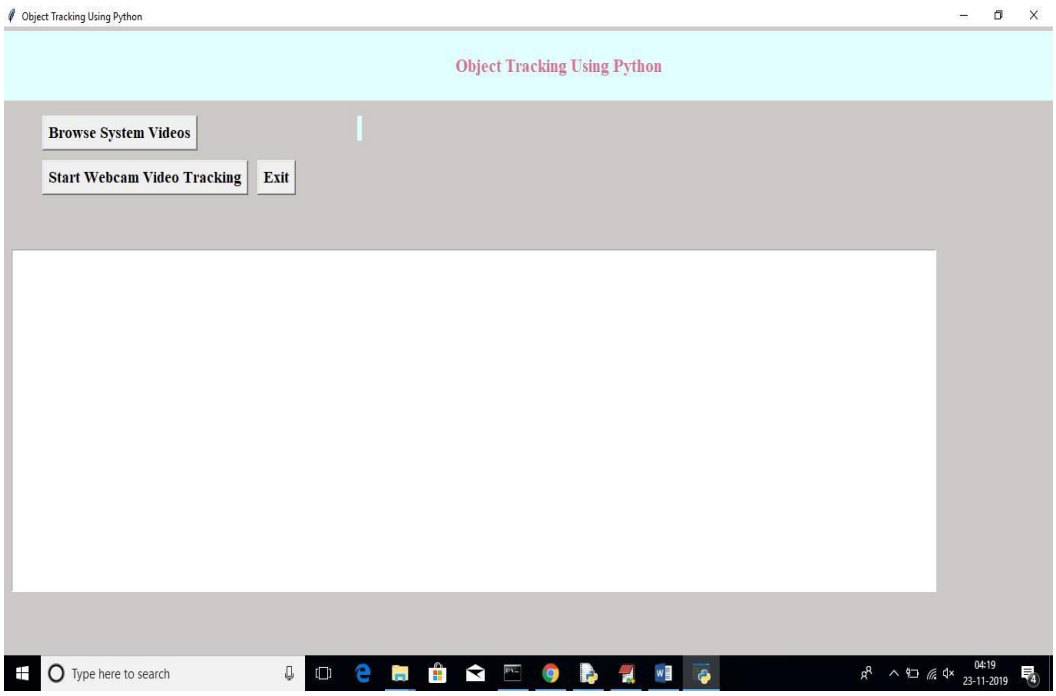
System Design



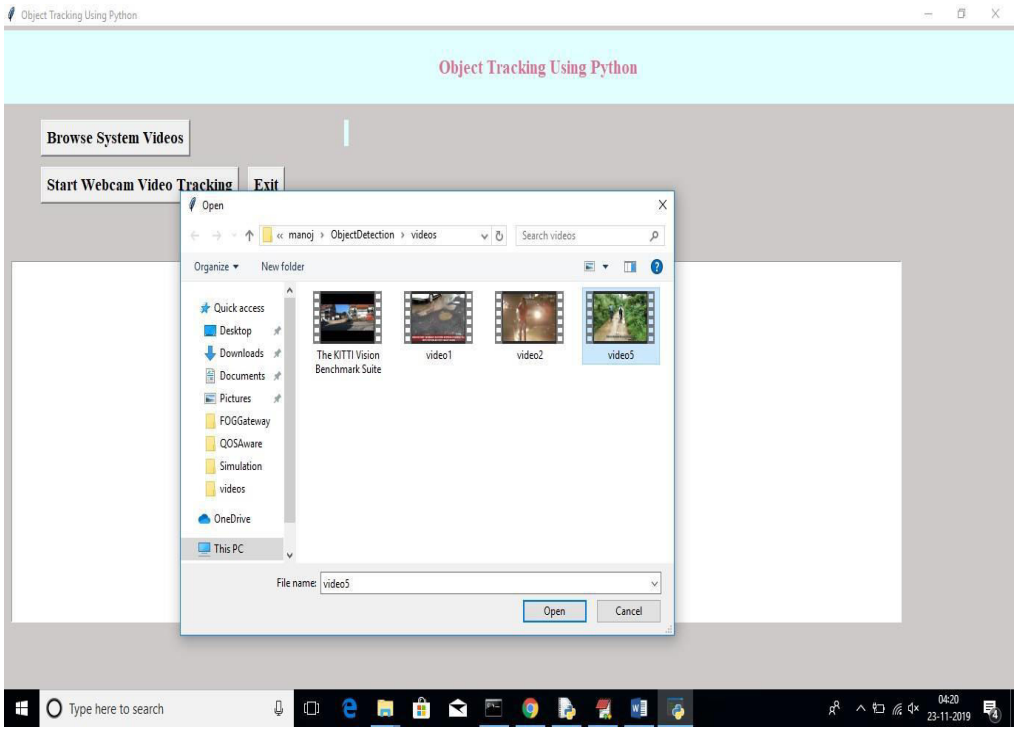
System Architecture

IMPLEMENTATION AND RESULTS

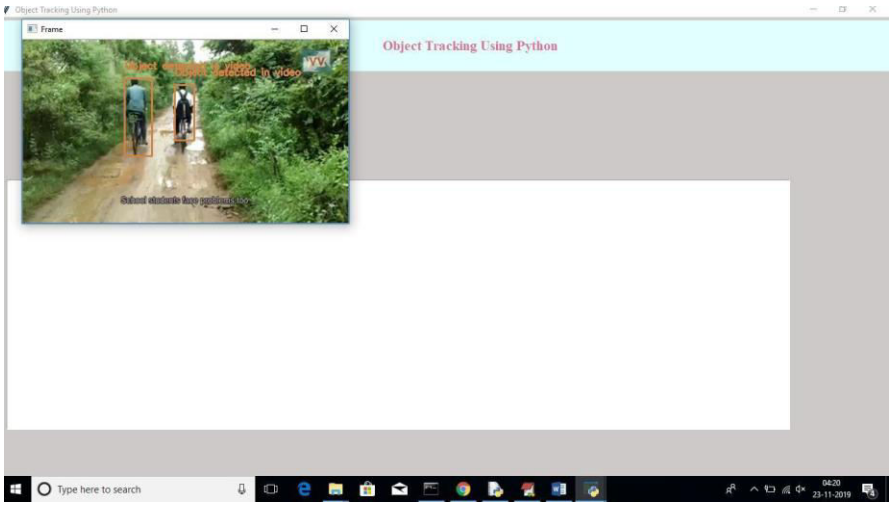
The given systems modules were implemented using python and the following shows the snapshots of the proposed systems output



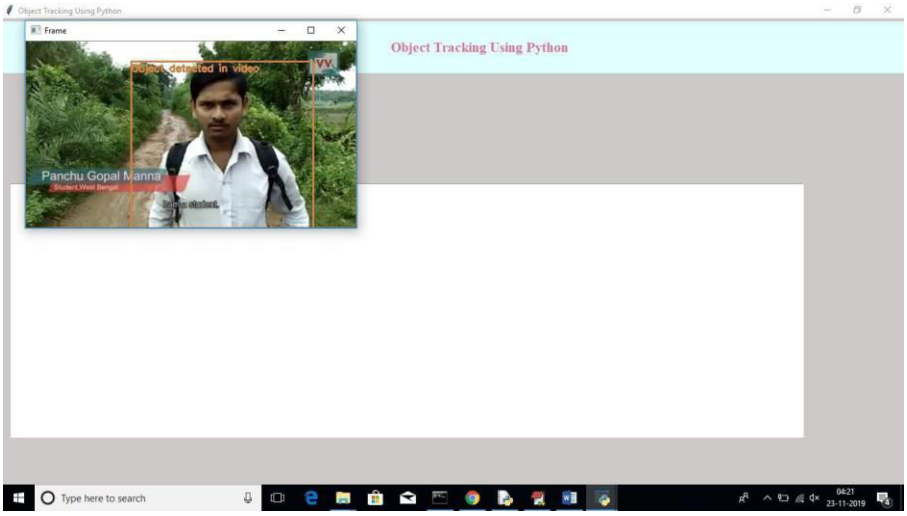
Now click on 'Browse System Videos' button to upload videos from system



Snapshot of Video Uploading Process



Tracking objects from video and mark them with bounding boxes.



CONCLUSION AND FUTURE WORK

A novel framework employing object detection, classification, and face and currency recognition has been presented to assist the visually impaired people. It is fairly simple, and easy to deploy, once the training part is complete. Using separate Inception models for faces and currency recognition makes it faster, user-specific and adaptable. It is one of the most generic framework, integrating all the useful features, and will surely prove to be a great service to mankind. Future work can be done 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] K. Bhargavi. An Effective Study on Data Science Approach to Cybercrime Underground Economy Data. Journal of Engineering, Computing and Architecture.2020;p.148.

[10] [21] M. Kiran Kumar , S. Jessica Saritha. AN EFFICIENT APPROACH TO QUERY REFORMULATION IN WEB SEARCH, International Journal of Research in Engineering and Technology. 2015;p.172

[11] K BALAKRISHNA,M NAGA SESHUDU,A SANDEEP. Providing Privacy for Numeric Range SQL Queries Using Two-Cloud Architecture. International Journal of Scientific Research and Review. 2018;p.39

[12] K BALA KRISHNA, M NAGASESHUDU. An Effective Way of Processing Big Data by Using Hierarchically Distributed Data Matrix. International Journal of Research.2019;p.1628

[13] P.Padma, Vadapalli Gopi,. Detection of Cyber anomaly Using Fuzzy Neural networks. Journal of Engineering Sciences.2020;p.48.

[14] Kiran Kumar, M., Kranthi Kumar, S., Kalpana, E., Srikanth, D., & Saikumar, K. (2022). A Novel Implementation of Linux Based Android Platform for Client and Server. In A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems (pp. 151-170). Springer, Cham.

[15] Kumar, M. Kiran, and Pankaj Kawad Kar. "A Study on Privacy Preserving in Big Data Mining Using Fuzzy Logic Approach." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 11.3 (2020): 2108-2116.

[16] M. Kiran Kumar and Dr. Pankaj Kawad Kar. “Implementation of Novel Association Rule Hiding Algorithm Using FLA with Privacy Preserving in Big Data Mining”. Design Engineering (2023): 15852-15862

[17] K. APARNA, G. MURALI. ANNOTATING SEARCH RESULTS FROM WEB DATABASE USING IN-TEXT PREFIX/SUFFIX ANNOTATOR, International Journal of Research in Engineering and Technology. 2015;p.16.