Deep Learning-Based Object Detection and Recognition Framework for the Visually Impaired (DL)

Dr. SMITA KHOND[1], K. RITWIKA[2], G. SRUJANA[3], G. JYOTHIRMAI[4], K. LAHARI[5]

[1]Associate professor, Department of Information Technology Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

[2,3,4,5]UG Students, Department of Information Technology Malla Reddy Engineering College for Women (UGC-Autonomous) Maisammaguda,Hyderabad-500100

ABSTRACT:

In this challenging evolution, the primary task in detecting the objects requires a computer vision that deals over indoor and outdoor classes. Over the past decades, this zeal requires more attentiveness. Previous implementation techniques involve in object detection with a strategy of single labelling. In this regard, a multi-label approach using machine learning and vision technologies, and accurate response can be acknowledged based on its accuracy and effectiveness. In the proposed work, we solve the existing system problem by using classification/clustering techniques that are used to reduce the recognize time of multi objects in less time with best time complexities. The model used to assist the visually impaired people can independently recognize objects which are near to them. The reverence, combined with the study, confounded the inception of these machine learning algorithms for visually impaired persons in assisting the accurate navigation, including indoor and outdoor circumstances.

Keywords: Deep Learning, Object detection, ML.

1. INTRODUCTION:

People with visual impairment face many difficulties in their lives. Recent data released by the World Health Organization (WHO) in 2020 shows that approximately 2.2 billion people worldwide suffer from imaginative and predictive disabilities. Detecting and identifying common devices in the area seems like a heroic project for the visually impaired. This is because they rely on different human beings; blind people rely on them and their sense of touch and smell to detect things, which can be particularly debilitating and sometimes dangerous.

White cane is the most popular blind navigation tool. It is further enhanced by including ultrasonic and IR sensors to remove obstructions inside the visually impaired person's space and detect vibrations or sound. Although this method proved useful for the movement of blind people, it provided very little information about the

environment. To better understand the user environment, it is important to know the objections and categories, along with identification and audio feedback.

Neural networks, especially CNN, have shown encouraging outcomes, especially in image object detection, type, and popularity. Most of the authors used the feed-forward neural network to offer sound cues about the purchased product. Implemented within the phone-based, real-time, complete disability detection and class gadget detection method, including the removal and monitoring the hobby element through a multi-scale Lucas-Kenned set of principles. Includes estimates of historical movement using homographic variations and a collection. The clustering method then uses classification, using the Histogram of Oriented Gradients (HOG) signifier in the bag of visual words (BoVW). Finally, a review of Electronic Travel Aids (ETA) constructed for visually compromised navigational aids is provided. Different ETAs, strengths, and weaknesses are mentioned and compared in terms of function. It also highlights that no state-of-the-art device has all the necessary features and that no generation needs to update the cane but completes it through appropriate warnings and comments.

OBJECTIVES

Specifically, the target consists of four subdivisions: DJCam for inserting the image into the frame, item identification and category module, face and currency identification module, and audio for the visually impaired user. In this paper, we concentrate on the framework's design, including article discovery, category, and reputation. SSD has proven to be the fastest single-image detection for multiple classes. The PASCAL VOC 2007 dataset contains 20 element classes, 9,963 images, and 24,640 annotated devices. This dataset has been changed by counting large snapshots inside the training, validation, and testing set, bringing the total training to 21, with the latest beauty being foreign money. The changed PASCAL VOC 2007 dataset is prepared by sharing knowledge with the SSD version.

PROBLEMSPECIFICATIONS

A new, deep novel architecture for the visually challenged, using long pauses of two parallel CNNs, outperforms Kingdom of Artworks' strategies for celebrity activism. CNN's GoogleNet and Alex Net pair each other by discovering unusual characteristics of the same glory, because of which they are each fed an input video feed and a support vector machine (SVM). The output beauty values are matched using.

2. LITERATURE SURVEY

Few products are being developed using in-deep learning and image processing to help the visually impaired. Below are some of the more common related jobs in this section. An Android mobile application was developed for the visually impaired, where live feeds from the camera were fed as input to a trained system learning model. The version uses image processing and item detection to capture instance tools that have been used to train versions on object classification. The software aims to help blind people move around without difficulty and alert them if they encounter any obstacles. A similar prototype tool has been developed, equipped with imaginative and present binocular sensors. These binocular sensors capture images at a set frequency from which the information maxima are determined by stereo image quality assessment (SIQA). These snapshots are sent to the cloud for similar processing. Detection and automatic results can be provided using a convolutional neural network-based primarily on large records. Through image analysis, cloud computing will return the requested information to the user so that the user can make cheap decisions about further actions. A mobile alert system has been developed for the visually impaired with a state-of-the-art and creative work-tracking infrastructure that uses web centers in the network to capture it. Mobile devices and marked points are determined by GPS function. The purpose of the cellular device is to enable visually impaired people to navigate distances according to the marked points. This research aims to explore and estimate the distance and relative functions of blind people around the devices around them, especially parked bicycles. The proposed tool uses a single-shot Multi-Box Detector (SSD) to locate parked bicycles using various learning algorithms. It is based on the principle of a triangle comparing the original motorcycle with the motorcycle image to estimate the distance of 2 to 5 meters of the motorcycle using a hollow chamber device. The study provides relevant information, including length and grip protection on flat floors. Pipelines combine multi-factor cloud instances, tabletop aircraft detection, device detection, and full version estimation through a robust estimator. In this bodywork, the benefits of recent deep insights (e.g., RCNN, YOLO) are used, which will be a green path for the detection project, while the geometry-based strategy estimates the full 3D version. -Relevant heritage objects no longer need to be separated (or separated) from the surrounding landscape for this study.

Mona and Riyadh proposed their dissertation in 2020 on "Retina Net Assessment for Internal Object Discovery for Assisted

Navigation for the Blind and Visually Impaired." In these articles, the feature of Computer Vision is to accurately locate objects inside the house. People with visual impairments can be helped by visiting the features of the CNN Framework.

To identify specific devices first, we want to find the pixels found in the images. Unfortunately, if the conditions of the lighting fixture are incorrect, it isn't easy to locate and locate the fixture with high accuracy. Next, the algorithm needs to extract the photographic capabilities with a selective spectrum to detect internal objects, which can be implemented through the retina net. Enabling small object detection networks through the Region Proposal Network (RPN) means sub-sampling to capture photographic records. With 152 samples, the resort achieved an average accuracy of 83.1%, and Dennis Net achieved an average accuracy of 79.8% with 121 samples.

Han Hu and Jiayuan proposed their paintings in 2018 on "Relationship Networks for Object Detection." Based on the relationship models, the work allocated the same amount of work, considering its characteristics. It eliminates duplication and achieves the exact requirements. Since objects are connected on a 2D scale, use elements instead of phrases. In addition, versions are categorized as extras below the geometric and actual weight. Xiangrong and Allen proposed their paintings in 2005 on "A one-time waterfall for real-time object detection: with applications for the visually impaired." In this work, the main objective is to identify the complexities of time and its accuracy based on it. Various tests are performed by the module capture system that detects textual content in images that can advance for the visually impaired. The beauty of the model can be measured by F.P. And F.N. quotes. The functionality of the algorithm's choice can be implemented through a set of classifiers and schooling images. It can use smart telescopic systems for people with vision problems. In micro-screen images, the image forcefully represents itself, leaving behind the positive aspects of the image.

Alice Tang and Xiwan proposed their work in 2018 on "Serial Cerebral Angiography Automated Registration: A Comparative Review." Over time, performed fine-tuning to identify and reverse the disorder, primarily based on this particular work in the medical field. First, consider its effectiveness and accuracy. Magnetic resonance imaging (MRI) and computed tomography (C.T.) scans are analyzed in an image processing algorithm that can better evaluate than the DSA. Although the DSA establishes a diagnosis of several neurovascular conditions used during surgical procedures, based on these considerations, we can conclude that the

system was designed based on patients diagnosed with ischemic stroke. Is. Wei and Xia proposed their articles in 2015 on "HCP: A Flexible Convolutional Neural Network (CNN) Framework for Multi-Label Image Classification." In this article, a version of CNN offers the best performance for image classification with a single tag. Due to its complexity, multiple labeling is an open plan for educational image design. The same image is taken as input to make assumptions, and it is shared with CNN to rank people through maximum polling. Single colors evaluate image hypotheses that different groups can represent.

The extraction technique produces predictive effects that can use to maximize polling. When reviewing I-FT and HCP models, the HCP version improves device performance by 5.7%. Rum and Assam launched their paintings in 2018 in a combination of "RGB (red-green-blue) image and related photography intensity (RGBD) images" of a complex network of complex values for the blind. This work uses multiple versions for visually impaired people to face objects more spectacularly in an interior area. This model has more than one label at a time. CVNN and multi-label techniques combine the image with labels that instantly match objects classes.

EXISTING SYSTEM

Low vision or blindness is one of people's top ten most common disabilities. Unfortunately, India is home to one of the most visually impaired populations in the world. This paper provided a unique framework to help visually impaired people find and recognize objects and is a great way to move around freely and keep track of their surroundings. The report uses a broadcast to provide insight into unpaired snapshot detection (SSD) approaches to locate and classify objects that can be examined when human faces and banknotes are detected. , Inception was completed using the v3 form. The entire SSD detector is based on the modified PASCAL VOC 2007 dataset, which introduces a completely new feature for coin detection. In addition, the discrete models of Inception v3 can recognize human faces and banknotes, making the framework scalable and adaptable to a person's ability. Ultimately, it can give the frame a visually impaired position within the sound layout. That changed the Delivery Forex Elegance (MAP) SSD independent detector rating to 67.8%, and the accuracy and forex credentials of the Inception v3 male and female models were 92.5% and 90.2%, respectively.

Proposed system:

In this mission, we combine SSD300 (Single Shot Detector) with the Starter version to search and evaluate currency banknotes. SSD 21

can recognize lessons, but it will not recognize currency bills, so we have introduced a broader layer for finding currency bills, but its accuracy is not sufficient. We removed the capabilities from the SSD and then returned to training with INCEPTIONV3. , Which can achieve accuracy. More than 97%. To find the currency, we used the INDIAN OLD NOTES dataset because the new Forex Note dataset is unavailable, so we train the SSD and start with the old Forex note, which this new model Can usually find 21 classes, of which 1 is more excellent. Forex, so now SSD General Configuration 22 can detect and identify.

3. METHODOLOGY

An in-depth, novel framework for the visually impaired using an ancient fusion of parallel CNNs that outperforms more sophisticated techniques for the popularity of hobbies. Both Google Net and CNN's Alex Net complement each other to identify unique features of the same magnificence. As a result, each of them is fed an input video feed and a support vector. magnificence beauty ratings are combined using a machine (MVS). Another proposed new method using CNN consists of a recurrent neural community (RNN) and a SoftMax classifier for detail detection and shade thresholding for hue, saturation, and intensity (HSI) of color reputation. . An approach that blends innovative and predictable wearable strategies with the profound talent of the visually impaired. The device uses regression-based mechanisms to scan items without preferred dates, handles sudden camera movements, and uses You Only Look Once (YOLO) to perceive the item.

A mobile application designed for blind men and women. You can draw in different ways: online and offline, based on the community connection between people. The online mode uses the fastest RCNN to make predictions in strong situations and YOLO for instant results. However, using the Haar and Histogram of Gradient (HOG) functions in a distinct identity module serves this purpose in the offline mode. CNN is designed to use the ImageNet dataset for the reliability of pre-qualified articles. A new DLSNF (usually a deep learning-based sensory navigation framework) based on the YOLO framework is proposed to develop a sensory navigation device based on the NVIDIA Jetson TX2. SqeezeNet, a moderate CNN pre-qualified model, scored higher and lower arithmetic delays. Squeeze Net was developed by changing the weight of the last convolution layer, replacing the corrected linear unit (ReLU) as an activation function with LeakyReLU, and adding a batch normalization layer.

OPERATION:

Object Detection Using Python

In this project using python, yolov3 (You Only Look Once v3) and OpenCV algorithms we are detecting objects from video and images. Yolov3 is a famous object detection algorithm developed by Washington university, this algorithm generate yolov3 weight model using python Deep Learning algorithm called CNN (Convolution Neural Networks). This algorithm is pre-trained with all images and assign unique class name to each unique images and then generate a model, this algorithm convert each images into layers and then for each layer extract features and add weight to the model, due to all possible features from single image another image with some related features can also be predicted. Whenever we are giving new image then that image will be applied on pre-trained weight model to get best accuracy matching image label.

You are asking to detect object without using any pre-trained model and it is highly impossible as all Deep Learning CNN networks works by using pre-trained models only.

To run this project you need to use below commands

python yolo.py image images/test.jpg

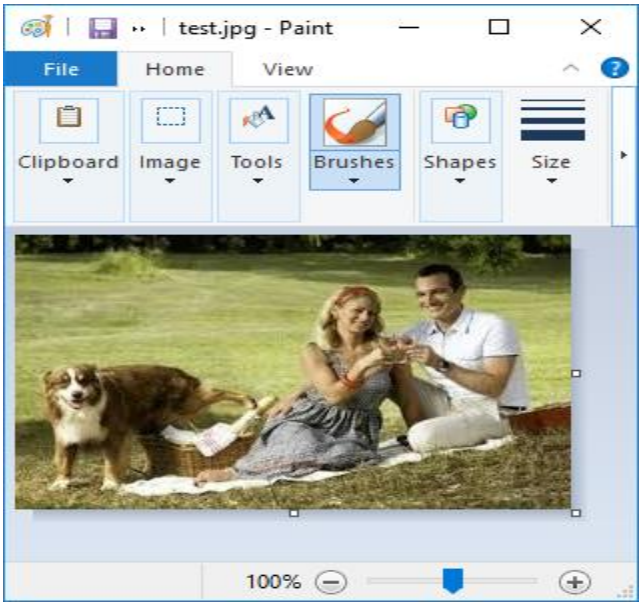
In above command python is the software name and yolo.py is the program name and 'image' means we want application to detect object from image and 'images/test.jpg' is the input image. Similarly for videos instead of image we need to pass video with video path.

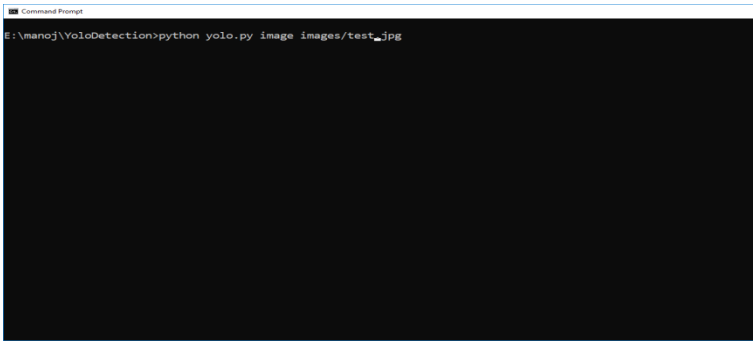
python yolo.py video images/video5.mp4

When we are giving video then it will take time to extract all frames from video for object detection and then create a new video called 'newvideo.avi' in the same code folder. If it's taking long time then you can press CTRL+C to stop execution and then you can play 'newvideo.avi' file.

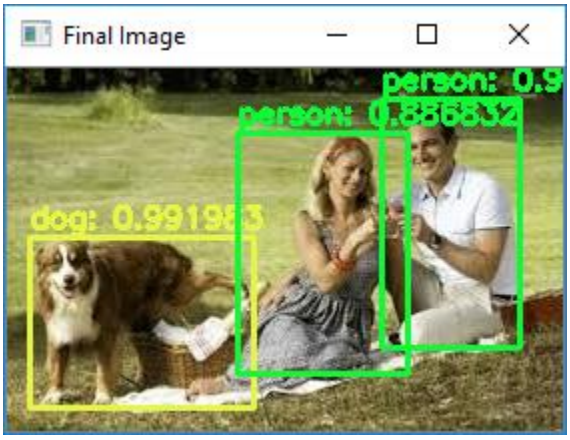
Screen shots

For testing I am using below image

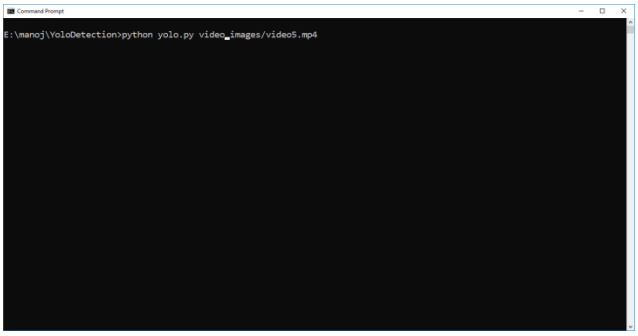




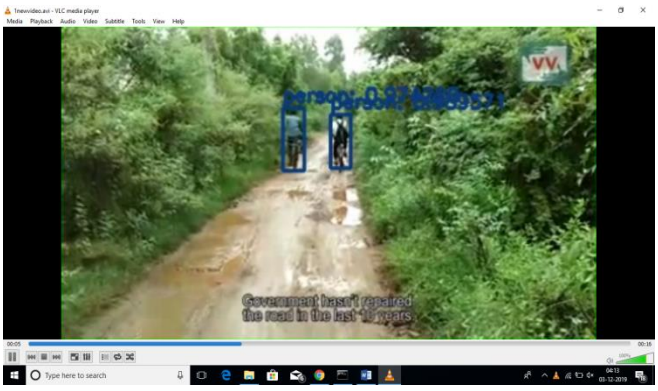
Using above command running yolo from image, after running above command will get below screen



Use below screen command to run with video



After running above command it will start generating new video when we play will get below screen



CONCLUSION

A new framework has been set up to help the visually impaired using item identification, type and face, and the popularity of Forex. Installation is relatively easy and clean after completing the training part. The use of separate start models for the popularity of faces and Forex makes it faster, more people-centered, and more adaptable. This is one of the most common frameworks that combines all the useful capabilities and can really be a great service to humanity. Future paintings can be completed to show the face and the popularity of Forex as spoof proof.

REFERANCES

[1] F. Jabeen, and A. M. Enriquez, 2015, "Feed forward neural network training based interactive shopping for blind,", IEEE, pp. 1–6.

[2] R. Tapu and T. Zaharia, 2013, "A smart phone-based obstacle detection and classification system for assisting visually impaired people,", pp. 444–451.

[3] R. Tapu, and E. Tapu, 2014, "A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments,", IEEE, pp. 1–4.

[4] J. Monteiro, R. Granada, 2017, "Virtual guide dog: An application to support visually-impaired people through deep convolutional neural networks,", IEEE, pp. 2267–2274.

[5] R. Kumar and S. Meher, 2015, "A novel method for visually impaired using object recognition,", IEEE, pp. 0772–0776.

[6] Tapu R, Mocanu B, 2017, "Seeing without sight-an automatic cognition system dedicated to blind and visually impaired people,", pp. 1452–1459.

[7] B.-S. Lin, and P.-Y. Chiang, 2017, "Simple smartphone-based guiding system for visually impaired people", p. 1371, 2017.

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

[9] J.-C. Yang L, 2018, "A deep learning approach to sensory navigation device for blind guidance,", IEEE, 2018, pp. 1195–1200.

[10] Y. Bazi and H. Alhichri, 2019, "Helping the visually impaired see via image multi-labeling based on squeezenet cnn,", p. 4656, 2019.