IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⋙ Impact Factor 7.047 ⋙ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10742

OBJECT RECOGNITION AND LOCALIZING For VISUALLY IMPAIRED PEOPLE

Vishwas B M[1], M N Chandan[2]

PG Scholar, Dept. Of MCA, P.E.S College of Engineering, Mandya, Karnataka, India[1]

Assistant Professor, Dept. Of MCA, P.E.S College of Engineering, Mandya, Karnataka, India[2]

Abstract: The number of people with visual impairments is considered to be in the hundreds of billions on a global scale. Integration into society is a crucial and continuing objective. Substantial work that has gone into ensuring a health care system. The creation of numerous guide system techniques has made it possible for people who are blind to live more regularly. These systems often only consider a limited set of functions. However, these solutions can greatly determine the movement and protection of such people.

Keywords: Visually impaired, YOLO, CNN, Recognition, Localization

I. INTRODUCTION

Modern approaches to image processing and computer vision, as well as the speed performance of the devices and unit processors, are all important in the development of cutting-edge navigation systems to help persons who are blind. Regardless of the technology used, the application must run in real time with prompt actions and judgments because action taking may depend on speed. Choosing the optimum solution essentially involves making a trade-off between the hardware capabilities and the software component's performance. The best parameter tweaking is necessary. One of the primary goals of the aided system is to automatically identify and distinguish objects or impediments during an indoor movement of a visually impaired individual, followed by an auditory alert. The platform created to assist visually impaired persons is integrated with the vision module for image processing that is suggested in this system. Additionally, the suggested module can be used separately from the integrated platform, outside of the shell. The proposed vision-based navigation system was created, tested in trials, and then iteratedly improved. The module adheres to the idea of creating a high-performance device that is also affordable and useful. The module uses innovative technology that enables updating and the addition of new features.

II. LITERATURE SURVEY

The portable system includes external sensing modules in addition to being built around a smartphone. It includes visually challenged people's movements both indoors and outside. Tests conducted demonstrate the system's effectiveness, which may be increased with the creation of portable android-based devices. This research introduces a portable solution to help the blind in both indoor and outdoor settings. It makes use of a variety of sensors to find impediments and utilises GPS and a compass to direct them as they go. An Android smartphone with many cores makes up the system's major component. Other sensory modules locate obstructions and provide pertinent data to the central component. For optional remote monitoring, the system can interact [1]. People with visual impairment encounter a variety of difficulties in their everyday lives since current assistive technology frequently falls short of consumer expectations in terms of cost and amount of aid. This research introduces a brand-new style of assistive smart glasses for blind pupils. The goal is to help with many daily chores by utilising the benefits of wearable design. This study only provides one example application—text recognition software that facilitates reading from hardcopy materials—as a proof of concept. The Raspberry Pi 2 single board computer serves as the brains of processing, while the Raspberry Pi 2 camera captures images, lowering the cost of construction. Results from experiments show that the prototype is operating as expected. [2]. This study explains a CNN-based correlation method to help persons with visual impairments. Whatever the variation, a visual processing unit must be incorporated into the design of systems that help persons who have vision impairments since so much information may be gleaned from the pictures that are gathered. In order to improve the characteristics of aiding systems and provide visually impaired people with additional information about their surroundings, this research offers a correlation method based on the usage of cellular neural networks (CNNs). The majority of the operations (calculations) in the suggested approach can be completed by parallel processing. Thus, it can speed up computation and the speed of computation won't grow correspondingly as the size of the template photos increases. [3].

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⋇ Impact Factor 7.047 ⋇ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10742

III. PROBLEM STATEMENT

A challenge in and of itself is coping with eyesight loss. Persons who really are blind or low eyesight are frequently kept in isolation by factors such as the lack of emotional support at diagnosis centres, the difficulties accessing activities and information, the stigma associated with the disease in society, and the dearth of employment options. A person who is blind or visually impaired could have trouble navigating a room or corridor without bumping into things. Obstacle avoidance can occasionally be challenging, uncomfortable, and possibly inaccurate, even with help, like a walking stick. Existing blind-guiding technologies, including Guide Canes and Smart Vision, use ultrasonic or laser sensors that emit and receive reflected waves to identify obstacles in front of the blind. When it notices obstructions, it emits an auditory signal or a vibration to alert the blind. Systems like Sound View take pictures with a single camera or two stereo video cameras installed on a wearable device. Scaled versions of these images are then further processed to create sound, audio, vibrations, or musical notes. These systems connect the pixel orientation and warning sound signal frequency. Some devices, like the Ultra Cane, use sensors to gather data and then provide advice to the user via vibration or sound. to support those who are blind.

IV. PROPOSED SYSTEM

By enabling more accurate object detection in a range of applications, such as general medical help and blind navigation equipment, deep learning systems have improved device performance. Systems are available to satisfy different needs, and a number of them do a good job at managing the detection of static obstructions. To evaluate the effectiveness of the indicated course of action, six pilot tests were carried out with successful outcomes. The dataset comprises with over 1000 categories, and a deep convolution neural network (CNN) and YOLO model are used to detect and classify objects with an accuracy of 90.3%. A quantitative comparison study based on the scores is also performed using the allowed functionalities of a devices. The suggested system has demonstrated to be more effective than the existing ones, with a total score of 9.1/10, which would be 8% higher than the second-best devices.

V. METHODOLOGY

A camera is used to record the video, which is then separated into a series of frames. Message to audio conversion is carried out using pyttsx3, and object detection is carried out using CNN classifiers.

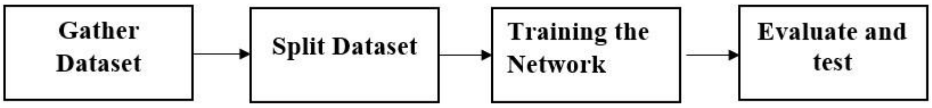

Fig 1: Deep Learning Steps

The cycle of image acquisition, image processing, and audio notification is repeated throughout the person's whole journey inside the building. The acquire rates for the supplied image frames is determined by adding the three processing times together, which results in the total processing time. The procedure must go quickly enough to allow for timely avoidance of any potential obstacles.

IMPLEMENTATION

The Python model loads all necessary libraries before starting to record video frames from the camera. The frames go through pre-processing such frame resizing, image enhancement, and blob format conversion. Once the pre trained models have been loaded, they process the input blob image via the layers to identify and classify the item. After classifying the object, the model plots the bounding box, determines its location, then extracts the object name and supplies it to the Pyttsx3 library to provide speech output.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526
International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.047 ⁜ Vol. 10, Issue 7, July 2022
DOI: 10.17148/IJIREEICE.2022.10742

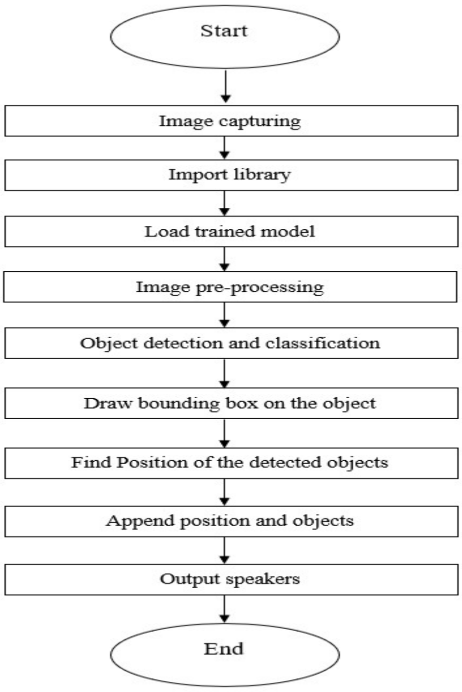

Fig 2: Flow of Implementation

VI. ALGORITHMS USED

A. YOLO

You Only Look Once is known by the term YOLO. This algorithm identifies and finds different things in a picture (in real-time). The class probabilities of the discovered photos are provided by the object identification process in YOLO, which is carried out as a regression problem.

The YOLO approach uses convolutional neural networks (CNN) to instantly detect objects.

1. Residual blocks
2. Bounding box regression
3. Intersection Over Union (IOU)

Residual blocks: The image is initially divided into a number of grids. The size of each grid is S x S. The following graphic shows the grids that were produced using an input image. There are numerous grid cells with equal dimensions in the figure below. Every grid cell will be able to detect items that enter it. For instance, a grid cell will be in charge of detecting an object if its centre appears within that cell.

Fig 3: Residual Boxes.

Bounding box regression: A bounding box is an outline that draws focus to a certain object in a photograph.Each bounding box in the picture possesses the qualities listed below:

- ➢ Width (bw)

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

ISO 3297:2007 Certified ⁂ Impact Factor 7.047 ⁂ Vol. 10, Issue 7, July 2022

DOI: 10.17148/IJIREEICE.2022.10742

- Height (bh)
- Class (for example, Traffic-light, car, person, etc) - This is represented by the letter c.
- Bounding box center (bx,by)

YOLO use a single bounding box classification to ascertain an item's height, width, centroid, and category. The figure below shows how likely it is that an object will be found within the bounding box.

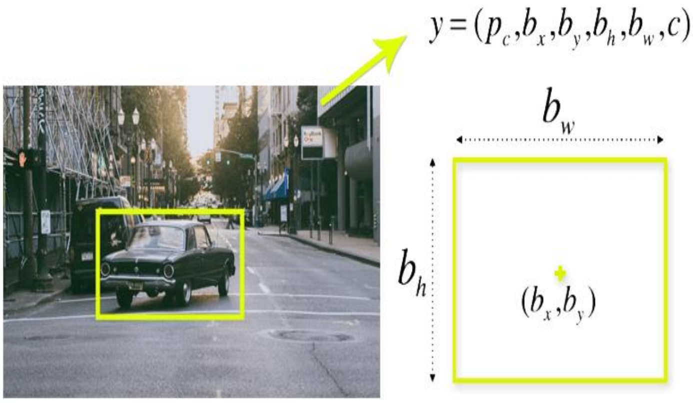

Fig 4: Bounding Box.

Intersection over union (IOU): Box overlapping is described by the discovery of objects phenomena known as intersection over union (IOU). IOU is used by YOLO to create an output box that properly encircles the items. Each grid cell is accountable for the anticipated bounding boxes and their confidence scores. The IOU is equal to 1 if the projected and actual bounding boxes match. This method eliminates bounding boxes that are not equal to the actual box.

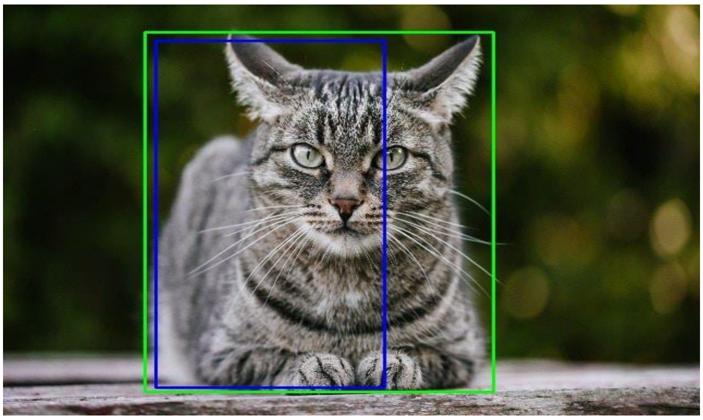

Fig 5: Intersection over Union

Two bounding boxes, one in green and the other in blue, are seen in the image above. The green box is the actual box, while the blue box is the anticipated box. The two bounding boxes are balanced thanks to YOLO.

B. Convolution Neural Network (CNN)

Deep Neural Networks' subset of convolutional neural networks (CNNs), are frequently used for visual image analysis because they can recognise and categorise specific features from images. Their uses include picture and video recognition, image classification, and many other things.

DOI: 10.17148/IJIREEICE.2022.10742

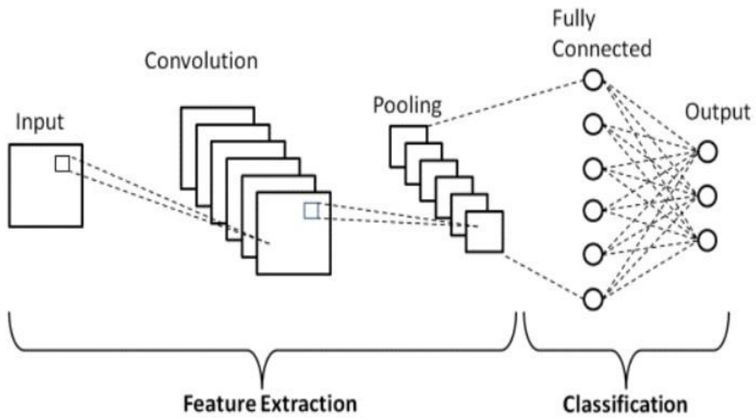

Fig 6: CNN Architecture.

A CNN architecture consists of two fundamental components:
Feature extraction is a procedure used by a convolution tool to separate and identify the distinct characteristics of a picture for analysis. A fully connected layer that makes use of the convolutional process's output and determines the class of the image using the features that were previously extracted.
Convolutional Layer: This is the initial layer that was used to extract the various characteristics from the input photographs. At this layer, a mathematical operation called convolution is performed between the input picture and a filter with the dimensions MxM. This gives information about the picture, including its corners and edges, and is known as the feature map. Later, additional layers are given access to this feature map so they may learn new features from the input picture.
Pooling Layer: After a convolutional layer, a pooling layer is frequently used. The primary objective of this layer is to reduce the size of the convolved feature map in order to reduce computing costs. Using fewer linkages between layers and separately modifying each feature map, this is accomplished. Typically, the Pooling Layer acts as a link between the FC Layer and the Convolutional Layer.
Fully Connected Layer: The neurons between two layers are connected using the Fully Connected (FC), which also contains a set of weights. These layers make up the final few of a CNN architecture and are generally located far before output layer. This gives the FC layer with a flattened version of the input image from the layers below. The flattened vector is subsequently subjected to the standard mathematical processes through a few further FC levels. The categorization procedure starts to take place at this point.

CONCLUSION

A vision-based guidance module will be made available to assist those who are visually impaired with the use of the YOLO library, which may be utilised to construct a a multiresolution strategy and outstanding outcomes in indoors applications. How precisely the vision module works might be upgraded by taking into account Adjusting the attribute to varying lighting circumstances for a legit-world situation.

REFERENCES

[1] P. Szolgay,L. Ţepelea, V. Tiponuţ, A. Gacsádi, "Multicore Portable System for Assisting Visually Impaired People", 14th International Workshop on Cellular Nanoscale Networks and their Applications, pp. 1-2, University of Notre Dame, USA, July 29-31, 2014.
[2] E.A. Hassan, T.B. Tang, "Smart Glasses for the Visually Impaired People", 15th International Conference on Computers Helping People with Special Needs (ICCHP), pp. 579-582, Linz, Austria, 2016
[3] L. Ţepelea, A. Gacsádi, I. Gavriluţ, V. Tiponuţ, "A CNN Based Correlation Algorithm to Assist Visually Impaired Persons", IEEE Proceedings of the International Symposium on Signals Circuits and Systems (ISSCS 2011), pp.169-172, Iasi, Romania,2011.