AUTO-ASSISTANCE SYSTEM FOR VISUALLY IMPAIRED PEOPLE USING DEEP LEARNING

G. Durvasi[1,] G. Divija[2], D. Venkata Maheshwari[3], V. Meghana Sai Sree[4]

Assistant Professor, B. Tech Students
[*] Department of Information Technology
[**] ANDHRA LOYOLA INSTITUTE OF ENGINEERING & TECHNOLOGY

Abstract- World Health Organization (WHO) outlined that there are two eighty-five million visually disabled worldwide. Among them thirty-nine million people are completely blind. One of the difficult activities that could be conducted by visually impaired is object detection which could be implemented using Machine Learning. It is an approach towards Artificial Intelligence that provides system the capacity for natural learning and development from experience without specifically programmed. It provides computer vision to the system which make decisions based on training algorithms. The chief goal of this research paper is to develop an object detection system to assist totally blind individual to manage their activities independently. Paper also compares different object detection algorithms like Haar Cascade and Convolutional Neural Network with yolov5(CNN). Haar Cascade classifier is a basic face detection algorithm which could also be trained to detect different objects whereas convolutional neural network falls under deep learning approach which could be employed for object recognition. The custom dataset is created with 2300 images consisting of 3 different classes. This comparison is being executed to find the yolov5 as a suitable algorithm for this system from the aspect of accuracy for real time scenario.

Keywords- Object Detection, Computer Vision, Deep learning, Feature Extraction and Recognition , Convolution neural network.

I. INTRODUCTION

Object detection and classification is a continuously examined field in computer vision. It enables computer and computing devices that mimic human eyes for object detection and perform computational processing. Computer vision (CV) becomes a novel inclination towards research, for building AI networks. Object Detection and classification is a key part of computer vision which consists of different Machine Learning and deep learning algorithms. Network is trained using a training data set for creating a model. On the basis of the model, prediction is made when new testing data is introduced to the ML algorithm, while with accuracy predicted input data is evaluated and deployed if found relevant. On the other hand if found not relevant it is trained iteratively again. Deep learning is subset of artificial intelligence that acts like human brain for handling data and generating patterns which helps in decision making. Deep learning makes use of a hierarchical level of artificial neural networks to carry out the task of machine learning. Many applications like human-computer interaction (computer responding/interacting with humans after analyzing what human feels), computer forensics (in the case of lie detection), pain detection, the field of education (i.e. distance learning where teachers determine whether the student understood the course), games and entertainment (for asserting user experience) find its base in facial expression recognition systems.

II. AIM & OBJECTIVE

To implement yolov5 + CNN + LSTM for captioning Images from captioning the scene and giving output as voice.

We analyze the scene from the image using yolov5 and detecting the objects with the help of LSTM we are trying to convert the image scene to LSTM.

III. EXISTING SYSTEM & ITS LIMITATIONS

In the Existing System was reference base papers they are using the same technique but they have created an android application.

LIMITATIONS

- Visually impaired people cannot see anything but how they will use the application.
- In the entire reference paper they haven't mention about the voice output so we created a model which will give output as voice for captions generated

IV. PROPOSED SYSTEM & IT'S ADVANTAGES

Python is the most preferred programming language for developing machine learning models and rapid app development because of its simple and concise codes, extensive selection of libraries and frameworks and platform independence. OpenCV is a computer vision and machine learning. Software library used for image processing. OpenCV is the most preferred library for Real-time applications because of its improved computational efficiency and distinguished set of libraries. We are using It consists of three parts: (1) Backbone: CSPDarknet (2) Neck: PANet (3) Head: Yolo Layer. The data are first input to

CSPDarknet for feature extraction, and then fed to PANet for feature fusion. Finally, Yolo Layer outputs detection results (class, score, location, size). LSTMs have an edge over conventional feed-forward neural networks and RNN in many ways.
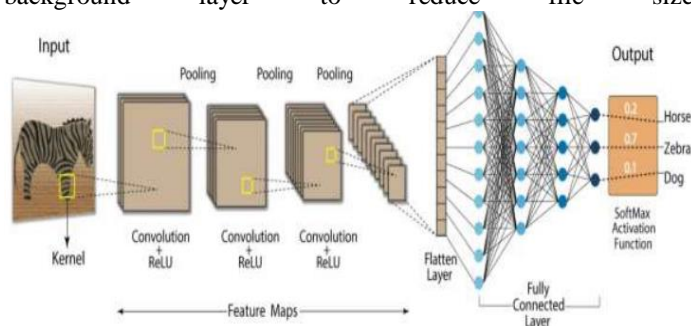
ADVANTAGES

- In this proposed system explain are the first attempt to detect the object.
- And then generate the captions with the generated objects

IV.   STUDY OF THE SYSTEM

CNN & MAXPOOLING

Convolutional layers are the layers where filters are applied to the original image, or to other feature maps in a deep CNN. This is where most of the user-specified parameters are in the network. The most important parameters are the number of kernels and the size of the kernels. Pooling layers are used to reduce the dimensions of the feature maps. Thus, it reduces number of parameters to learn and the amount of computation performed in the network. The pooling layer summarizes the features present in a region of the feature map generated by a convolution layer. Flattening is merging all visible layers into the background layer to reduce file size.

MODULES

There are 3 modules we are using in this project.
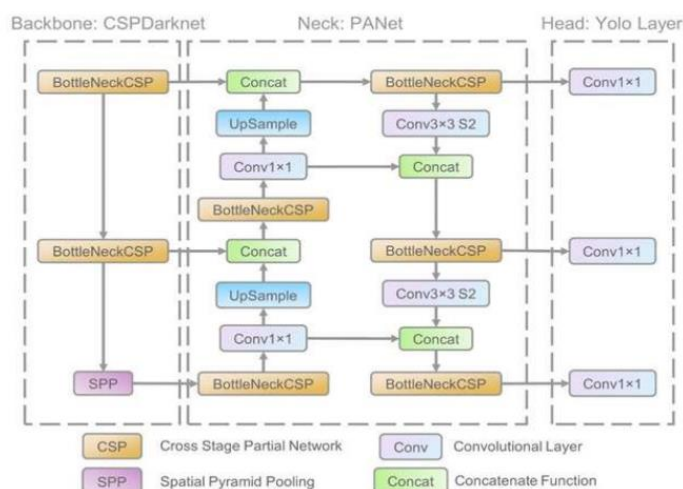They are: 1) User Activity 2) Training Dataset 3) Text Vectorizer
1. User Activity: User login our system or mobile phone or any text devices. Us clarify questions and response the admin every time in same questions but not answer the same answer. The admin replies for the user.

2.Training Dataset: We have to create the captions and objects for every dataset and we have to split the dataset for training, testing and splitting.

3.Text Vectorizer: We have to convert the words from training data to vectorizer It transforms a batch of strings (one example = one string) into either a list of token indices (one example = 1D tensor of integer token indices) or a dense representation (one example = 1D tensor of float values representing data about the example's tokens). This layer is meant to handle natural language inputs. To handle simple string inputs (categorical strings or pre-tokenized strings)

YOLO5

YOLOV5 is controversial due to the fact that no paper has been published yet (till the time of writing this) by its author Glenn Jocher for the community to peer review its benchmark. Neither it is seen to have implemented any novel techniques to claim itself as the next version of YOLO. Instead, it is considered as the PyTorch extension of YOLOv3 and a marketing strategy by Ultranytics to ride on the popularity of the YOLO family of object detection models. But one should note that when YOLOv3 was created, Glenn Jocher (creator of YOLOv5) contributed to it by providing the implementation of mosaic data augmentation and genetic algorithm. In digital photography, computer-generated imagery, and colorimetry, a gray scale image is one in which the value of each pixel is a single sample representing only an amount of light; that is, it carries only intensity information. Gray scale images, a kind of black-and-white or grey monochrome, are composed exclusively of shades of grey

Matplotlib:

Matplotlib can be used in Python scripts, the Python and I Python shells, the Jupiter notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. We can able to generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery. For simple plotting the pilot module provides a MATLAB-like interface, particularly when combined with I Python. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object-oriented interface or via a set of functions familiar to MATLAB users.

TensorFlow:

TensorFlow offers multiple levels of abstraction so you can choose the right one for your needs. Build and train models by using the high-level Keras API, which makes getting started with TensorFlow and machine learning easy. For more flexibility, eager execution allows for immediate iteration and intuitive debugging. For large ML training tasks, use the Distribution Strategy API for distributed training on different hardware configurations without changing the model definition.

Torch:

Porch is an open-source machine learning framework based on the Torch library, used for applications such as computer vision and natural language processing, primarily developed by Facebook's AI Research lab. It is free and open-source software released under the Modified BSD license.

V.   SYSTEM ARCHITECTURE

The complete working procedure of the project.

VI.   SYSTEM DESIGN

System design shows the overall design of system. In this section we discuss in detail the design aspects of the system.

VII.METHODOLOGY INVOLVED IN THIS PROJECT

Object Oriented Analysis and Design:

When Object orientation is used in analysis as well as design, the boundary between OOA and OOD is blurred. This is particularly true in methods that combine analysis and design. One reason for this blurring is the similarity of basic constructs (i.e. objects and classes) that are used in OOA and OOD. Through there is no agreement about what parts of the object-oriented development process belongs

to analysis and what parts to design, there is some general agreement about the domains of the two activities. The fundamental difference between OOA and OOD is that the former models the problem domain, leading to an understanding and specification of the problem, while the latter models the solution to the problem. connected rectangles in which the value of the feature is the difference of sum of pixel values in black and white regions.

Python Namespace:

A namespace is a way of providing the unique name for each object in Python. Everything in Python is an object, i.e., a variable or a method. In other words, it is a collection of the defined symbolic names along with the information about the object that each name references. Generally speaking, a namespace (sometimes also called a context) is a naming system for making names unique to avoid ambiguity. Everybody knows a name spacing system from daily life, i.e. the naming of people in first name and family name (surname). An example is a network: each network device (workstation, server, printer, ...) needs a unique name and address. Yet another example is the directory structure of file systems. The same file name can be used in different directories, the files can be uniquely accessed via the pathnames.
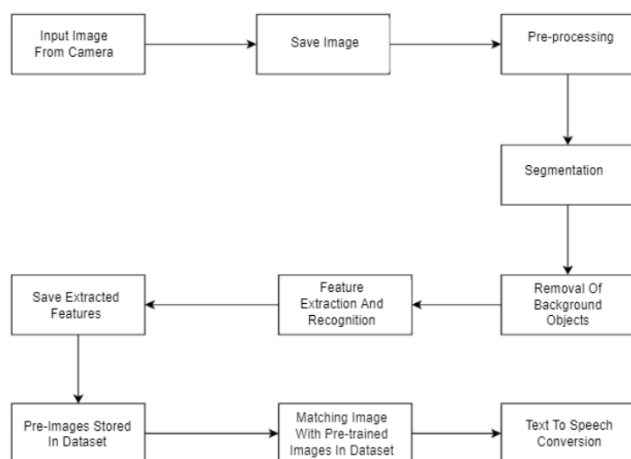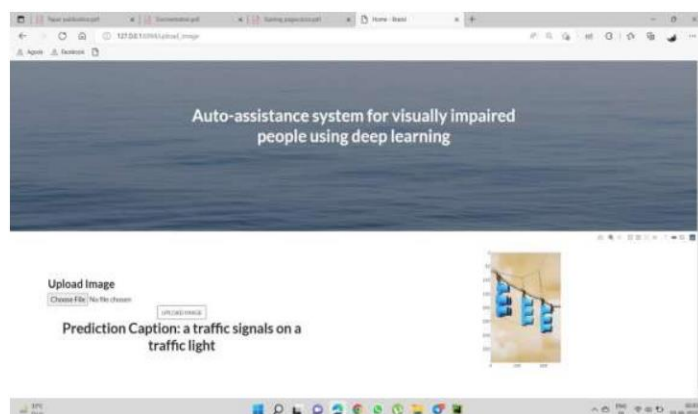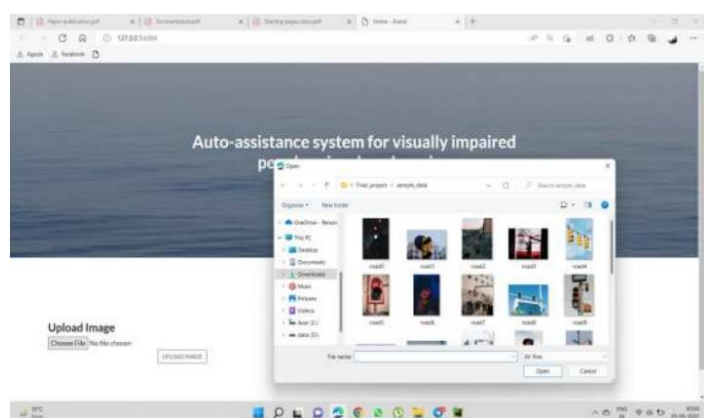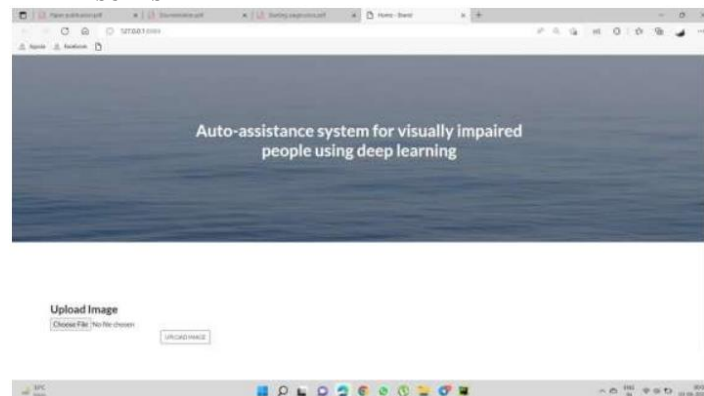
Python Web Frameworks

A web framework is a code library that makes a developer's life easier when building reliable, scalable and maintainable web applications. Web frameworks encapsulate what developers have learned over the past twenty years while programming sites and applications for the web. Frameworks make it easier to reuse code for common HTTP operations and to structure projects so other developers with knowledge of the framework can quickly build and maintain the application. Common web framework functionality: Frameworks provide functionality in their code or through extensions to perform common operations required to run web applications. These common operations include:

1. URL routing

2. HTML, XML, JSON, and other output format templating

3. Database manipulation

4. Security against Cross-site request forgery (CSRF) and other attacks

5. Session storage and retrieval

THE RESULTS

REQUIREMENTS

SOFTWARE REQUIREMENTS

OS: Windows or Linux

Python Version: python 3.9.12 and above

Setup tools and pip to be installed for 3.6 and above

Language: Python

HARDWARE REQUIREMENTS

RAM: 8GB and Higher

Processor: Intel i5 and above

Hard Disk: 10GB Minimum

VIII. CONCLUSION

Object detection algorithms are compared to detect multiple objects simultaneously. COCO 2017 Dataset has been used to train CNN model. When processed on CPU, CNN is the fastest algorithm, but YOLOV5 gives more accurate results when detecting multiple objects simultaneously for real time applications. Accuracy around 80% was found using CNN with the particular performance input parameters. GPU is the basic requirement to train dataset with a greater number of classes which leads to less computational time and high accuracy. A device with text-to-speech conversion could be used to provide information. The system for visually impaired could be a real boon to make the visually impaired person self-reliant and independent.

FUTURE SCOPE

In the present project we have seen how to implement the model with camera and speaker. So in future they can directly get voice guidance from the spectacle itself. So that there is no extra tools, hardware to be installed manually

REFERENCES

- https://ieeexplore.ieee.org/document/8373566
- https://ieeexplore.ieee.org/document/8480413
- https://ieeexplore.ieee.org/document/9306804
- https://www.researchgate.net/publication/338253407_A_STUDY_ON_OBJECT_DETECTION
- https://www.researchgate.net/figure/gTTS-converting-text-to-speech_fig3_339548395
- https://www.sciencedirect.com/science/article/pii/S2352914820305621
- https://ieeexplore.ieee.org/document/683777
- https://ieeexplore.ieee.org/document/9220797

AUTHORS

G.Durvasi[1] M.Tech., Assistant Professor, Department of IT.

G. Divija[2] B.Tech, Andhra Loyola Institute of Engineering & Technology

D. Venkata Maheshwari[3] BTech,Andhra Loyola Institute of Engineering & Technology

V.MeghanaSaiSree[4] B.Tech, Andhra Loyola Institute of Engineering & Technology