Proceedings of the 7th International Conference on Intelligent Computing and Control Systems (ICICCS-2023)
IEEE Xplore Part Number: CFP23K74-ART; ISBN: 979-8-3503-9725-3

2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS) | 979-8-3503-9725-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICICCS56967.2023.10142625

Robot with Speech Recognition, Sentiment Analysis, Object Detection

Rubeena Rustum
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,Telangana, India
rubeena.rustum@gmail.com

Vaddepally Laxmi
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,Telangana, India
vaddepallylaxmi8@gmail.com

Siripuram Navya
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,Telangana, India
siripuramnavya07@gmail.com

G. Naveena
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,Telangana, India
naveenaguthi6@gmail.com

V. Sravanya
Department of Computer Science and Engineering
Gokaraju Rangaraju Institute of Engineering and Technology
Hyderabad,Telangana, India
lovelyjaanu37@gmail.com

Abstract

This study introduces a small companion Robot and with some of its interesting features. Sentiment Analysis by facial expressions and sentences, Object detection by YOLOv3 algorithm, Sentiment Analysis by which the robot can recognize all organisms face and detect the emotion by facial expressions by using CNN algorithm and cv2, speech recognition to recognize the sentiment of the sentences. The Robot is made by using various sensors and actuators like 8*8 LED matrix which displays the emotion on the screen, it can move forward and backward by the geared DC motors and are implemented features of the Robot by using various types of Machine Learning Algorithms.

Keywords – Companion Robot , Sentiment Analysis, Object Detection, Speech Recognition, cv2(Capturing Video using OpenCV), YOLOv3(You Only Look Once version 3), Convolution Neural Network, Sensors, Actuators.

I. INTRODUCTION

Robotics is about us. It is the discipline of emulating human lives, of wondering how it works. These Robots are the reflection of ourselves. One of the kind of Robots are the humanoid companion Robots which interacts with people, moves forward and backward, detects the sentiment like happy, sad, neutral, surprise, angry by speech and facial expression and the other hand it also detects the objects which are in the surroundings of the Robot.

II. RELATED WORK

The following are a few research works related to this project:

The authors of the study[1] proposed the YOLO algorithm for detecting the objects in an image. Input image, YOLO config file, pre-trained YOLO weights, text file containing class names are the required arguments.

The authors of the study[4] used the Raspberry pi for the speech recognition and an external microphone and USB to connect. The authors explained various models by which the speech recognition can be done like HMM(Hidden Markov Model), DTW(Dynamic Time Wrapping), Neural Networks(DNN-Deep Neural Networks and RNN-Recurrent Neural Networks).

The authors of study[6] proposed the optimised techniques of the sentiment analysis using machine learning algorithms. Study explaines the Naïve Bayes, BFTREE for sentiment prediction, OneR. The optimised way is done through taking the python libraries like NLTK, bs4 and the raw data s processed.

The authors of the paper[13] suggested the MAX7219 IC to use with 8*8 LED matrix and through this we can display the emotion on this display.

All these are the features which are implemented individually, now here is such a project where the Robots have all these features in addition it can interact and move forward and backward.

III. METHODOLOGY

A. Object Recognition:

Here the Robot recognizes the object by using the YOLOv3 algorithm. It is mainly used to detect the objects surrounded by it.

Working of YOLOv3:

The algorithm mainly used to detect the objects in the videos, live feed, images. And it mainly uses the deep Convolution Neural networks.

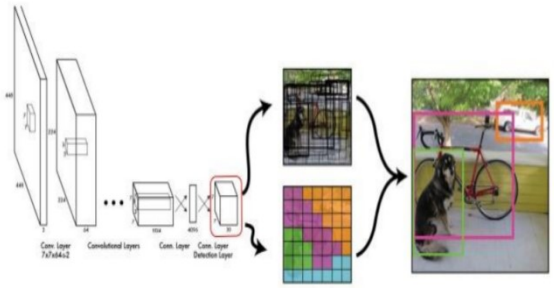

Fig1: Yol0V3 algorithm

B. Sentiment Analysis:

Here the use of sentimental expression like anger, cool, neutral, Happy, sad etc.. It also mainly describes the sentimental expressions of a person. These are quite useful for the people to know the mood of a person. The sentiment analysis is mainly used to identify the emotions that are associated with the humans and the robots. These also help us to identify the nature of the person and can act accordingly.

Fig2: Types of Sentiment

C. Speech recognition:

The speech recognition mainly identifies the technology and the methodologies that help to use the recognition and it also involves the translation of the spoken language into text format. It also identifies the spoken words and convert them into readable format.

Fig3 Speech Recognition

D. Hardware Requirements:

1. Arduino: It is a Microcontroller board which contains the ATmega328P microcontroller microchip. It contains analog and digital input/output pins.

2. Battery: The battery used here is 450mAh.

3. Speakers: The speakers used are of 3w.

4. Display: For the display of the emotions of the robot, 128*64 OLED and 8*8 LED Matrix are used.

5. Motor: Here, the various motors like Vibration motor ,Geared DC motor, Servo motor are used. These are used for the movement of the robot.

6. Sensors: Here various sensors like Touch sensors, Shock sensor, Infrared sensor are used. These sensors senses the related information and gives the outcome.

7. Speakers and IC 4301: Speakers and Integrated chip of 4301 and battery are used by which the Robot can speak.

8. Other Hardware used: Speakers, Buttons, Microphones, IC's.

IV .IMPLEMENTATION

The Django framework is used to display the software related results like CNN for sentiment, YOLOv3 for object recognition etc, and cardboard, sensors, actuators, IC's for the outer part of the Robot.

Different types of algorithms which are used here are:

A. CNN algorithm:

This algorithm is to detect the emotion or the sentiment by using the facial expressions.

CNN is one of the class that is being used in deep neural network. It is also one type of machine learning algorithm which takes the input image and assign the importance to various aspects and able to differentiate from one another.

The steps involved in the algorithm

1. Upload the Data set : The data set will be available with scikit can be download and can store them.
2. Input Layer: When the input layer is defined it is used to shape the data set and the first argument is also the feature of the data, that is defined.
3. Convolutional Layer: This layer helps to find whether the input and output both have equal padding or if not, it adds the zeros in rows and columns to ensure the same size.
4. Pooling Layer: This layer reduces the extension of data.
5. Dense Layer: It flattens the previous data to form fully joined layers and we can also use the different activation function and can dropout the data.
6. Logit Layer: The final step is this and here it predicts the data by using the concept called prediction.

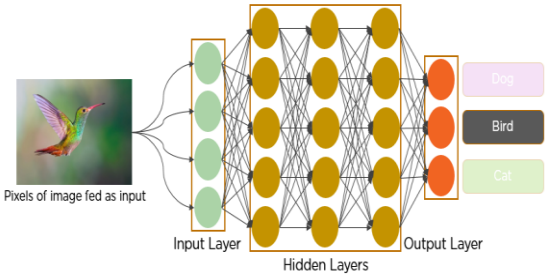

Fig4: CNN algorithm process

B. YOLOv3 algorithm:

It is the method used here to detect the objects, from the authors of the paper[1] image is taken, YOLO pre trained weights, it takes the different approach to get outcome from the object which it detects. It is very fast and most popular algorithm that is used for the object detection. In this, numpy, python, OpenCV dependencies that are useful for detecting. In this we use the command line arguments like input image, YOLO config file, Pre-trained YOLO weights, text file containing class names. By reading the input image one can get the width and height of image. The text file containing class names these are human readable format and can obtain a list from it. Here the use of multiple output layers for the prediction. One should check all the detection's from the output layer to get class id, confidence and bounding corners and can ignore the low confidence values from output layers. Thus it detects the objects precisely with confidence values. When we ignore the weak detection's there will be a lot of duplicates and they will overlap with the bounding corners. So here, the use of Non-max suppression which removes the high overlapping and display the output image.

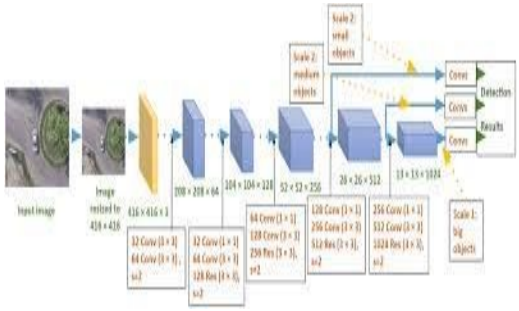

Fig5: YoloV3 Algorithm

Fig6: YOLO (You Only Look Once) Algorithm

C. Sentiment using audio algorithm: It shows the sentiments like positive, negative and neutral .It is used in application of machine learning and Deep learning since it gives the most accurate values precisely. In this the use of sentimental analysis on audio file will happen.

1. Firstly one should upload the audio file and then get the transcription of audio file
2. Now, perform the. Speech to text sentimental analysis and then it gives the output as positive, negative and neutral with the help of API
3. All the responses of the audio are stored in the text format and results are stored in a particular file called sentiment results and then

displays the emotions with the help of the responses that are stored in it.

The python library librosa provided the tools for processing and extracting the features from the audio files which are uploaded in the Django Framework.

From[17] after the foundation for this emotion detection using the audio file is feature selection. After this the functions of librosa library MFCCs, Chroma, Mel spectrograms are extracted from the audio files and then from the Sci-Kit Learn, the models are assembled.

Then we use the ensemble models which uses the soft voting classifier to improve performance and PCA(Principal Component Analysis) to extract features and these features which are extracted are used in the CNN(Convolutional Neural Network).

The Result will be on the form of any of the emotions such as Happy, sad, neutral, surprised, disgust and the accuracy will be shown. Figure 19 illustrates the emotion using the audio file when uploaded.

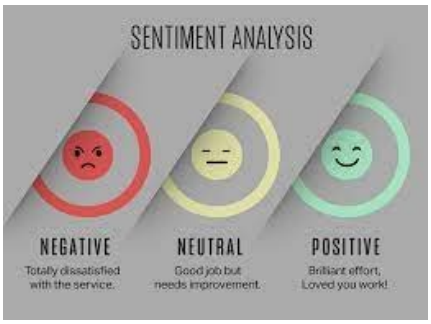

Fig7: Sentiment Analysis

D. Cv2: OpenCV is one of the library in python which is used to solve the computer vision troubles with it. It also reads the contents present in the file and if there are any missing files present in it ,It immediately displays the message on the output bar. It is mainly used for the images .And it also background color, color change and grayscale present in the image. And it displays the output in the compiler.

Fig8: A picture analysed using OpenCV

E. Sentiment Analysis using Live Camera

Here various training and testing data is used.

In training data contains various kinds of images with different resolution and various emotions. From paper[15] we can see that this proposed sentiment analysis contains three steps: training the model, feature extraction and predicting the result or testing. This proposed model detects or predicts the emotions such as happy, joy, sad, disgust, surprise, neutral.

Training data is given, the face extraction is transforming the raw data into numerical features. The training data sets when passed through prediction labels gives the classification, we use SVM here. Hence the sentiment or emotion can be classified by video, image or camera. In our project the emotion is predicted via Camera.

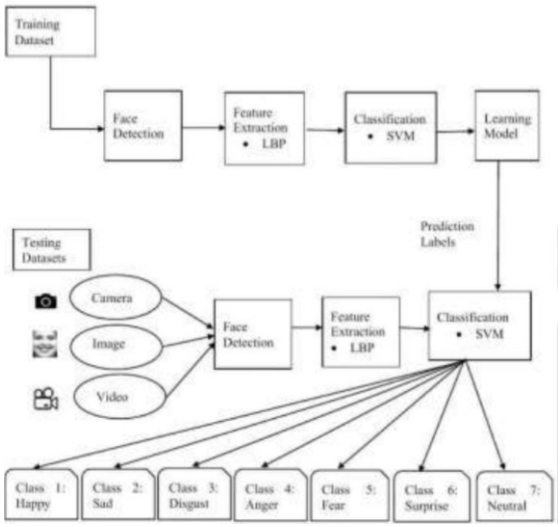

Fig9: Sentiment Analysis design

F. SENSORS

- Touch Sensors:

These the type of sensors which detect the touch. These are used in lamps, mobiles etc. These are simple and advance in technology. These sensors work similar with switch. The touch sensors are mainly used in this project since when we touch the sensor it automatically senses the touch and produce the light on it as the outcome which shows the touch sensor is enabled and proceed further.

Fig10: Touch Sensor

- Shock Sensor:

It mainly identifies the shock whether it is physical or the impact that has been occurred. It mostly detects the shocks present in the circuit and if any it identifies and gives the alert in the form of the sound or light form. For every material used here use have the shock sensor in it and it also controls the low voltage and high voltage present in the framing. In this project the use of shock sensors is to identify the defect of the wire and if any fluctuations are present .It mainly detects the shocks.

Fig11: Shock Sensor

- Infrared Sensor:

It is the sensor which emits the light in order to sense the objects of surroundings. It can measure the heat of the object and can also detect the motion of it. The radiations which are invisible to the eyes are detected by this sensor. It can also grab the information from satellite. In this project the use of this sensor is to detect the motion and heat of the object.

Fig12: Infrared Sensor

G. MOTORS

- Vibration Motor:

It is a three phase motor. It is used to alert the user from receiving the call by vibration or sound. It produces the centrifugal force where the high-offset from the shaft produces the more force and vibration amplitude .In this project we mainly used this motor in order to know the force and the speed produced in between the shafts. And the amount of force that is being used for the working of the motor more effectively.

Fig13: Vibration motor

- DC Motor: It is a motor which converts the direct electrical energy into mechanical energy. The motor speed can be controlled by using the variable supply voltage or by changing the strength of current. In this we use the motor to convert the electric energy into mechanical form.

Fig14: Geared DC Motor

- Servo Motor: This is a type of motor which control the circuit and provides the feedback of the current position of the shaft. It contains controlled device where it is controlled by the shaft and the motor, and the feedback is produced by it. In the project this motor to get the current position of the motor.

Fig15: Servo Moto

V. RESULTS

Robot is the machine resembling the human being. The intellect of the brain of the humans will be mimic by the Robots.

So we have designed a website so that we can depict the internal working of the Robot's features like sentiment analysis, object detection, speech recognition.

For the Robot, the hardware used are Arduino uno, and different types of the sensors and actuators example: motors. The below picture shows the outer part of the Robot:

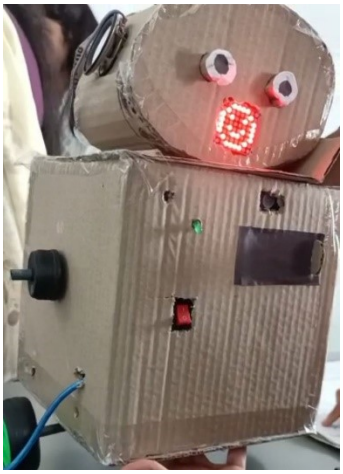

Fig16: Robot Showing its emotion (Happy)

Here in the above picture the Robot is in the happy state and the wheels at the bottom is used for the movement of it i.e to move front and backwards. We can also notice the speakers which makes the Robot to speak and can listen to it. This is possible by the IC used there. And Switch is used to activate that it can be switch on or off.

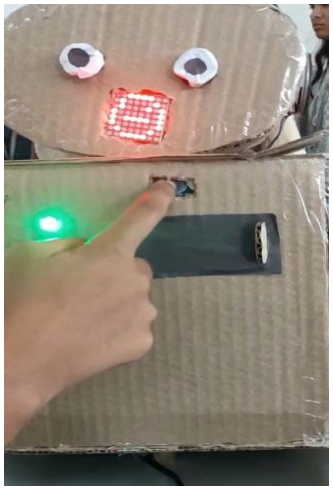

Fig17: The Led is blinking when one touch's it on the sensor

Here is the video link of the working of the Robot:

https://photos.app.goo.gl/nQLKm1WevFrJwQnz7

Video Link 1

The Django internal depicting of the Robot is:

The home page of the webpage look like

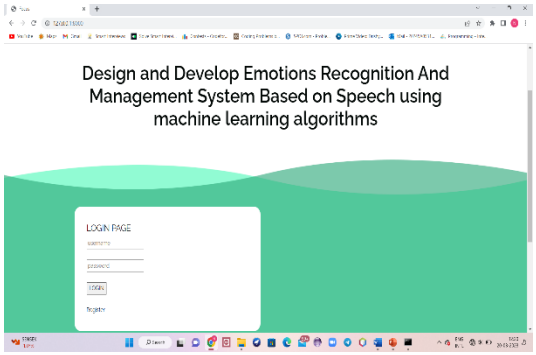

Fig18: Home Page showing the login

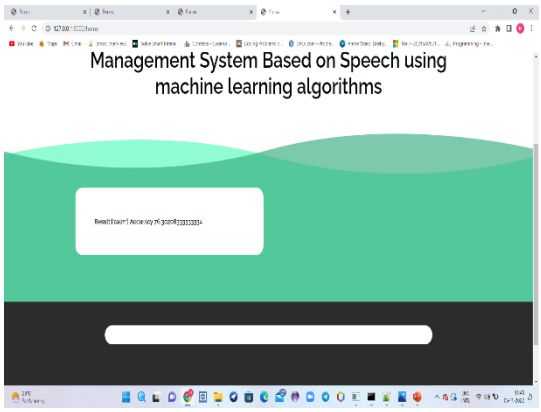

Fig19: Figure showing the accuracy of particular audio file based on the words and tone of sentences

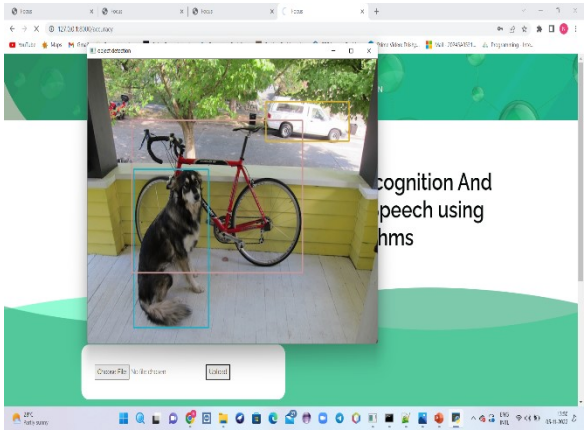

Fig20: Object Detection

The fig 20 is Showing the object detection when we upload any photo.

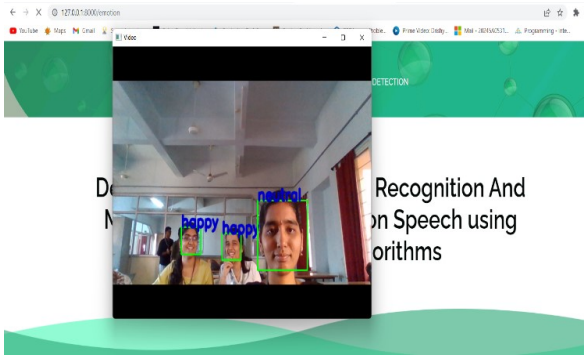

Fig21: Sentiment Analysis

The above figure21 is Detecting the emotion or sentiment using the facial expressions.

The training data when passed through the prediction labels classifies the data. The face will be detected and feature extractions takes place and then classifies the face accordingly to the emotions such as happy, sad, neutral, disgust, surprise. From[15] we can know that this process will be more accurate than the previous process of sentiment analysis.

This can be used in various purposes. Not only the Robot but in case of security also one can use this kind of projects to detect the humans.

VI. CONCLUSION AND FUTURE DEVELOPMENT

Hence from the above procedure we have seen that the Robot's hardware and software are merged to ensure the working of the Robot with human intelligence. The movements are shown and the internal working is depicted as a website.

The features of the Robot internally detects the objects by YOLOv3 algorithm which is the real time object detection algorithm that identifies the specific objects, another module is the emotion or sentiment by using the live video where the Robot look at the environment and detects the emotion of the persons around it, for this we have used training data consisting of thousands of images with different types of emotions, face detection, feature extraction, classification and finally predicting the emotion via live camera is the proposed procedure for sentiment analysis used in this project. One more feature of the Robot is recognizing the sentiment or the emotion by the audio. In this project we have used the python library librosa and funtions in it like Chroma , mel spectrograms.

All these features of the Robot are shown by using the Django Framework, Refer Figure 18,19,20,21 and the external parts of the Robot such as Geared DC motors help the Robot to move forward and backword, Touch Sensor Figure10 helps the Robot to sense when anyone touches it Figure17.

Refer Video link 1 for the Robot working Motion,voice etc.

Hence this Project of Robot can be a good companion to everyone in the family and this Robot not only used for as the Companion but also in various fields such as army, medicine etc.

The Future development of this Project will have more scope for this new digital era which is growing rapidly. The features can be implemented more accurately by using various other algorithms.

VII. REFERENCES

[1] M, Anmol. "YOLO Object Detection with OpenCV and Python." Towards Data Science, 29 Jan, 2021.

[2] Traversy, Brad. "Python Django Web Framework - Full Course for Beginners." YouTube, 11 May 2019.

[3] Aggarwal, V., and J. K. Sharma. "Machine Learning Paradigms for Speech Recognition: An Overview." ResearchGate, 2014.

[4] Kapila, Prashant, et al. "Experimental Speech Recognition System Based on Raspberry Pi 3." ResearchGate, 2017.

[5] Zhang, Jun, et al. "A Review of Audio-Based Emotion Recognition Techniques." Sensors, vol. 21, no. 16, 2021.

[6] Devi, B. Jayashree, and P. Dhavachelvan. "A Survey on Speech Emotion Recognition: Techniques and Datasets." Human-Centric Computing and Information Sciences, vol. 7, no. 1, 2017.

[7] Dey, Jayanta Kumar, et al. "A Review of Speech Emotion Recognition Techniques." IEEE Access, vol. 2, 2014.

[8] Shalini, M., and Sumathi. "How to Get Reports from Audio Files Using Speech Recognition and NLP." Towards Data Science, 8 Apr. 2021.

[9] Jha, Sudhanshu. "Sentiment Analysis: Assembly AI Python." Towards Data Science, 18 Mar. 2021.

[10] Shahbaz, Syed. "8X8 LED Matrix Interfacing with Arduino." Arduino Project Hub, 5 Feb. 2018.

[11] Bhatt, Devashish. "Control DC Motor Without IC - Motor Driver." DIYI0T, 23 Nov. 2021.

[12] Rui Santos. "Guide for OLED Display with Arduino." Random Nerd Tutorials, 3 Feb. 2022.

[13] Raj, Vikram. "8X8 LED Matrix MAX7219 Arduino." How2electronics, 2 Dec, 2021.

[14] Samuel Algherini, "How to get reports from audio files using speech recognition and NLP, Transform speech into knowledge with Hugging Face/Facebook AI and expert.ai", September 15, 2021

[15] Smita S Patil, Aishwarya A Patil, Shweta B Bamane, Sakshi S Karajgar, Snehal M Sankpal, Avinash Pratap Budaragade,"SENTIMENTAL ANALYSIS", ResearchGate, Article, February 2023.

[16] Md. Forhad Ali, Mehenag Khatun, Nakib Aman Turzo,"Facial Emotion Detection Using Neural Network", Research Gate Article, August 2020

[17] Ian Hatfield, Pedro Rivas, Tushar Gupta, Rocco Lange, and Arnav Deshwal. " Speech Emotion Detection, Extract Human Emotions from Audio Files", December 14, 2019.

[18]Konda Sai Varshitha, Chinni Guna Kumari, Muppala Hasvitha, Shaik Fiza, Amarendra K, Venubabu Rachapudi, "Natural Language Processing using Convolutional Neural Network", 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), pp.362-367, 2023.

[19]K.M.Rjesh M Naveen Kumar, "Artificial Intelligence Image Emotion Detection Mechanism Based On Python" International conference Electrical, Electronics, Communication, Computer and Opticational Techniques (ICEECCOT),1-5,2016