WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Scene Recognition Using Deep Learning

Vaibhav Bhalala[1], Hardik Vartak[2], Aradhya Parab[3], Dr. Archana Ekbote[4]

[1]Information Technology, Vidyavardhini's College of Engineering and Technology Vasai Road, India
[2]Information Technology, Vidyavardhini's College of Engineering and Technology, Vasai Road, India
[3]Information Technology Vidyavardhini's College of Engineering and Technology, Vasai Road, India
[4]Information Technology Vidyavardhini's College of Engineering and Technology, Vasai Road, India

ABSTRACT—

Scene Recognition is becoming a important Ap- plication of deep learning, with undergoing several important evolution over the past years. Scene recognition is used in so many fields from automated cars to indoor localization this proves the importance of scene recognition and its growth in future. This project is an attempt to solve one of the most common and difficult problems of scene recognition. Scene Recognition is a application with a aim to classify a given scene image to a predefined category by analysing the whole image. The process of classifying scene images is very difficult because of similarity in attributes of various scenes. The goal of this project is to create a cnn (Convolution Neural Network) model which can accurately detect a scene with the help of extracting features from the image and analysing it. We have used a custom dataset of 10 different classes to predict the scene and the accuracy of the model is 75 percentage.We will use Flask Framework for frontend to show output in realtime.

***Index Terms*—Scene Recognition, Deep learning, Convolutional Neural Network, Flask.**

I. INTRODUCTION

A. Overview

A human being can easily classify a object or a thing by looking at it because of the brain . The brain can analyse the object at a glance. To differentiate between two different things will be very easy and take mere seconds .In human brain there is a neural network and these objects act as a input for this neural network . But can a computer do this impossible task as sufficiently as a human person. Yes ,with the help of Artificial Intelligence and Deep learning. Scene recognition is a growing in the field a Deep learning and ML and the requirement of scene recognition in the market is high .Scene recognition is a very challenging issue that requires a systematic approach to get the needed results. Scene recognition is used in various fields like map construction, robots, AI automated cars and if there is some errors in making it can cause some serious damage . In This project we are going to use ANN architecture which stores information like a human neural network which can help in recognizing and classify a image.The ANN architecture is CNN(Convolutional Neural Network) which has a layered architecture for feature extraction and classification.

B. Scene Recognition using Deep Learning

For the past few years Scene Recognition has become very important and challenging issue in IT industries. Scene Recognition is used in many fields like security, indoor localization, Robots, autopilot. A scene consists of various concepts , including scene attributes, background, objects. These characteristics are used to determine the image. To extract these characteristics and train the model for detection we have created a cnn model. CNN (Convolution Neural Network) is one of the most useful method to for image classification . CNN is a type of Artificial Neural Network which is very effective in areas like image classification and recognition. In this project we have created a cnn model from scratch to train the large dataset and later to detect the images with respect to their class. [5]. When we have to recognize a coloured image or scene or a video, the CNN(Convolutional Neural Network) is the best and most suitable option because of it's architecture. The CNN architecture has three layers for feature extraction and two layers for classification. The feature extraction consist of an Input layer then the Convolutional Layer after that the padding layer which is optional and then the Pooling layer.In Classification layer the output from the convolutional layer goes into fully connected layer and finally the output layer.

C. Motivation

Identifying or classifying a scene through out eyes are human vision is easy but through computer is very interesting thing to work on as we see different scenes on our day to day life.Classifying and recognizing a scene through a computer is itself a challenging task.CNN is the best technique to do this task with utmost accuracy. That's the motivation behind this project. Big Data companies is a big market for scene recognition for ads preferences basis of majority of scenes detected of a particular user.

D. Problem Statement

In this project we have used a subset of a larger dataset of various images of different scenes from all over the places. The goal is identify the scene correctly. The cnn model we are going to make must be able to classify the scene correctly on the basis of training on the dataset helping in accurately identifying the image.

E. Organization and contribution of the report

This report summarizes the research and development ac- tivities carried out throughout the academic programme's final year, emphasizing the work conducted during the semester. Our efforts included feature extraction based on numerous website domain variables, model training, and requirement collection from diverse data sources.

II. LITERATURE SURVEY

Numerous researchers have created various methods for scene recognition for different purposes. Like sometimes for indoor localization or for military purposes. Some researchers have used more than one model or have used pre trained model for better accuracy. Different models like CNN and R-CNN and pre-trained models like FOSNet and ImageNet. Following is the outcomes of the literature survey that we performed for this project.

- In the research paper [1], the authors Junhyeok Lee and Bongjin Oh used two convolutional neural network where one cnn is used to train the big image dataset of different scenes while other cnn is used to extract the objects from that scene. They achieved a accuracy of 85 percent. While a hybrid of place365 and resNet achieved 87 percent.
- In this research paper [2], the author Haitao Zeng, Xin- hang Song and other senior members have approached the scene recognition problem by combining object features and scene features. they used two models one to extract the global features while other model to extract the local features and then aggregating them to achieve a new- state-of-art performance. They achieve a accuracy of 88 percent on MIT67 dataset.
- In this research paper [3], the authors Xie, Lin and Lee, Feifei and Liu, Li and Kotani, Koji and Chen, Qiu have created a Convolutional neural network to first extract prominent features from the images and then training the model and classifiying the images based on the features.

III. PROPOSED SYSTEM

A. General Methodology for Image Classification

Scene recognition through human vision is easier than through computer. The best method for a computers to un- derstand a scene and further classify and predict is through a ANN(Artificial Neural Network) model. ANN is a general model through which a computer can be train to classify and predict certain images. CNN(Convolutional Neural Network) is a type of ANN which specializes in image classification. Here is a general Block diagram of the CNN process. In CNN the model first reads the images from the training set which is mostly around 50000 images from the whole dataset. some pre-processing is done on the images so that all the images are of same size. Then the images are run through the cnn model for the training of the model. After training the model the model is used to test the model on test images of the dataset and then the model is saved for the predicting the class of a given image.

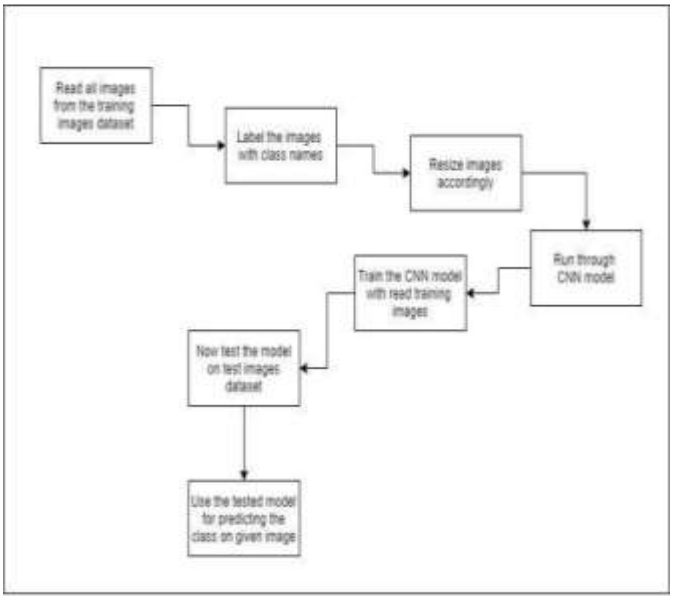

Read all images from the training images dataset
Label the images with class names
Resize images accordingly
Run through CNN model
Train the CNN model with read training images
Now test the model on test images dataset
Use the tested model for predicting the class on given image

Fig. 1. General Block Diagram

B. Hardware/Software required

Creating a CNN model is very difficult if you don’t have the required Hardware and Software for it. CNN requires a high computational power to train a model if the datasize is very big. The machine has to have high powering GPU and ram and high processing power of i7, ram of 12 to 16GB.For model building you need CNN , Tensorflow. To display the prediction you can use Flask in frontend. Jupyter Notebook to write the code and kaggle to get the dataset or you can make your own dataset.

C. Dataset

For this Project we have used a custom dataset of 10 dif- ferent classes like [Airfield, Aquarium, Supermarket, Desert, Waterfall, Bookstore, Landfill, Classroom, Train, Rainforest] where each class containing of 5000 images from all over the world, which is divided in training and testing .The dimensions of the images are 150x150. In future we are going to use a bigger dataset (places365 or SUN397) which contains 400 plus different categories with 40000 images for training .This will help for better accuracy of the model and better and vast prediction.

IV. METHODOLOGY

A. Image Pre-Procssing

Pre-Processing is a very important step in cnn model. Pre- processing takes 50 to 70 percent work of the whole project. Images comes from different sources in different format and size. Therefore Image pre-processing is very important other- wise it will be very difficult to train the model. The first task is to make the size of all the images same and convert the image in RGB channel.Now the next task is Data augmentation.

B. Data Augmentation

To avoid overfitting and increase the dataset’s variety, vari- ous data augmentation approaches were used to the training set using the Image Data Generator method of the Keras library in Python. The first task is to rescale the images i.e Normalisation

. Normalisation means rescaling the images so that they can lie in a confined range. Now the data augmentation part where the image is converted in to many images to increase the data size. This is used when the data size is very low , this helps in increasing the dataset with different variety to like performing zooming, croping,rotating the image, shear etc and finally creating a batch size for training and testing of 64 images with 3 distint channels.

C. Training, Testing and Validation

After Data Pre-Processing and Data augmentation now is the time to make a cnn model. DESIGN of cnn Architecture. For creating the model we have used a kernel size in Convolutional layer (3,3) with strides=0 with the filter=64.[4] In MaxPooling layer we have used the poolsize of (2,2) with strides =2, and finally in Dense(Fully connected layer) we have used the Relu activation function.The cnn has two important parts the feature extraction part and the classifier part.

In CNN we have three layers the First is the Input layer. Input layer is responsible to give the first convolutional layer the images. The second layer in cnn is the Convolutional layer.

In convolutional layer can have multiple layers from 2 to 16 or even more.The Hidden Layer consists of convolutional layer,pooling layer, padding layer.

1. The Convolutional Layer: Convolutional means a math- ematical operation which involves to function to produce a third function. In cnn convolution is performed on the input data with a filter to produce a feature map.Here filter is kernel of size(3,3) and feature map is the dot product of filter and input data.

2. The Pooling Layer: The Pooling layer is added after the convolutional layer for continuous dimensionality reduction i.e to reduce the parameters thereby reducing the computational time and controlling overfitting. Pooling can be done in two ways ,Max pooling and average pooling. We have used Max pooling which takes the maximum value in each window which reduces the feature map size while keeping the relevant information.

3. The Dropout:A technique where we randomly selected neurons are ignored during the training phase.This is a method which is used to avoid overfitting the model and to get well generalized result.

After this we flatten the inputs from ND to 1D and it does not affect the batch size and last is the dense layer which is a fully connected with it’s preceding layer. In dense layer activaltion function is performed which is used for the transformation of the input values of neurons. We have used Relu Activation Function. ReLU function works as if the input is equal or greater than the threshold value than the output will be the input value otherwise it will be zero. After this the model is fitted and the accuracy of the model is checked. We got the accuracy of 75 percentage, then the model is saved for prediction.

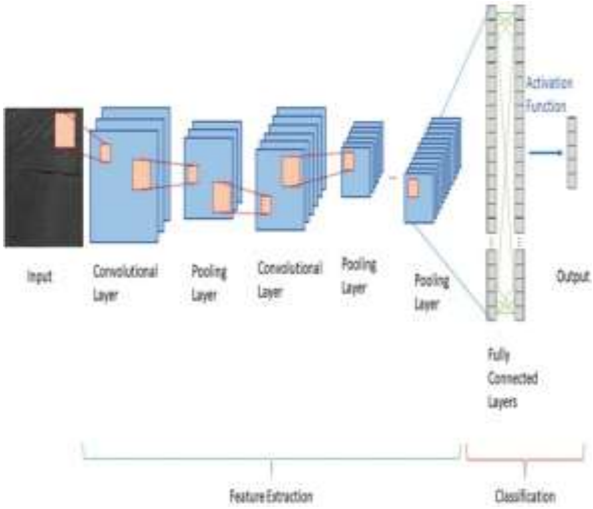


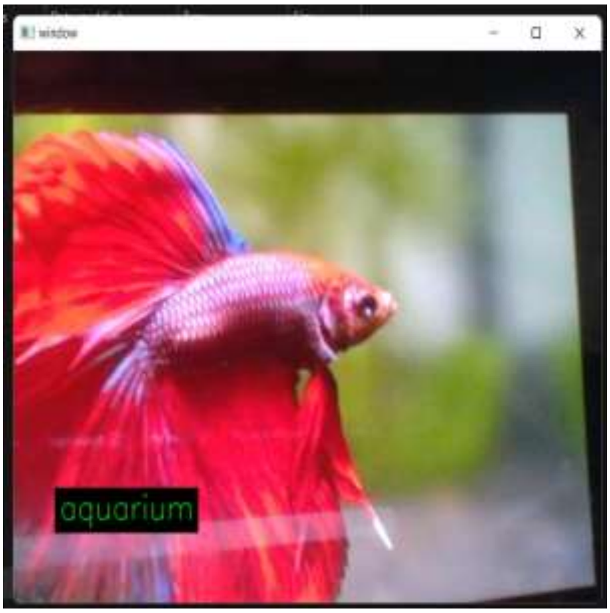
Fig. 2. General Block Diagram

V. RESULT AND DISCUSSION

For this Project CNN model has been trained to predict and classify a image in 10 categories with each class has around 50000 images in Training set from all over the world and 1000 for testing with image size of 150x150.The images are in RGB. In the dataset there is a folder with images which are not in any particular class so that it can be used for the prediction purpose. The images also has RGB channels with image size of 150*150.On Jupyter Notebook, the proposed model experiments were carried out. We can predict scene by uploading images of different scenes or in real-time as well. We have created the frontend using flask.

1) Running Prediction on Samples: Below are the outputs of the predictions that we performed on a sample.

We have done prediction on both images as well as real-time using CV2. On some data we have got the correct prediction will on some we have got no prediction or wrong prediction.



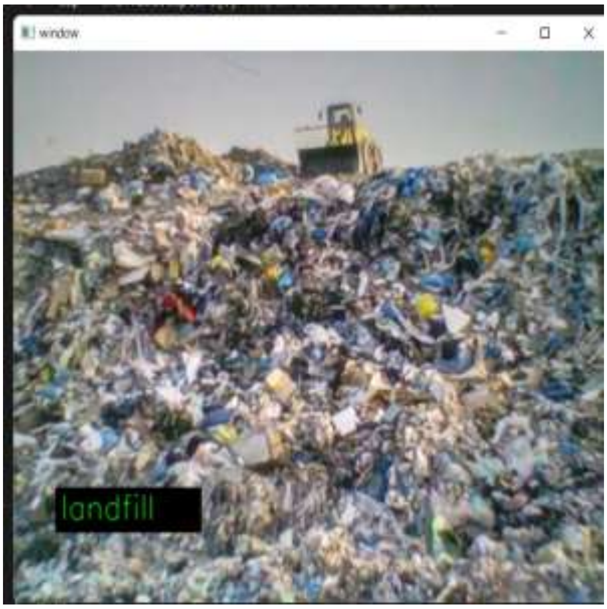


Fig. 3. Prediction of Aquarium Scene

Fig. 5. Prediction of Landfill Scene

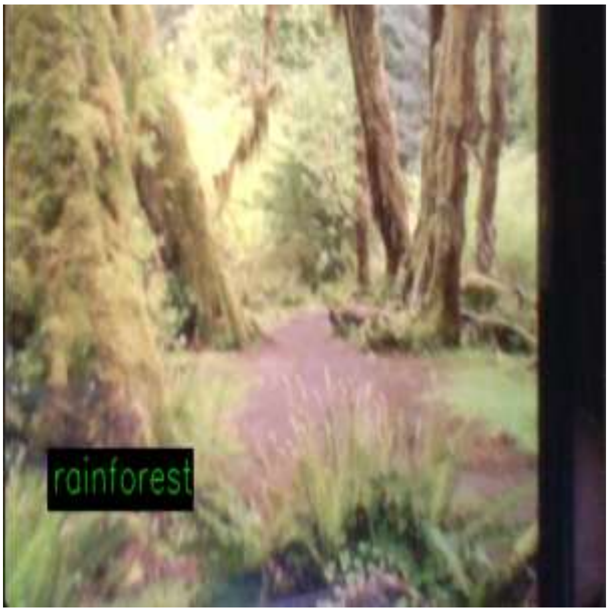


Fig. 4. Prediction of Rainforest Scene

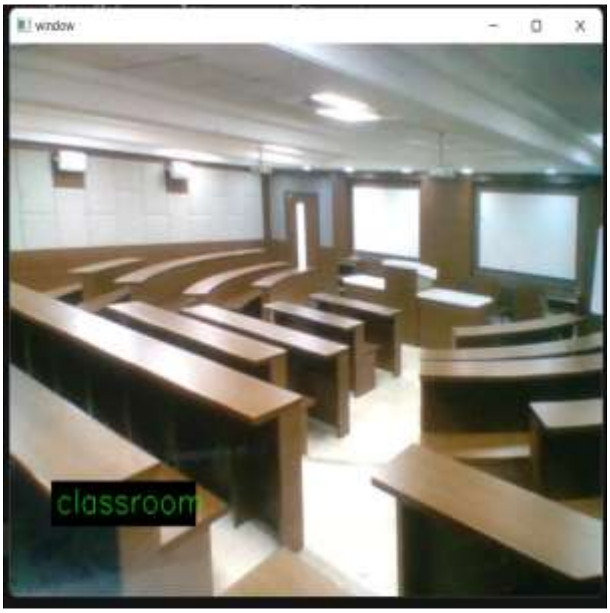


Fig. 6. Prediction of Classroom Scene

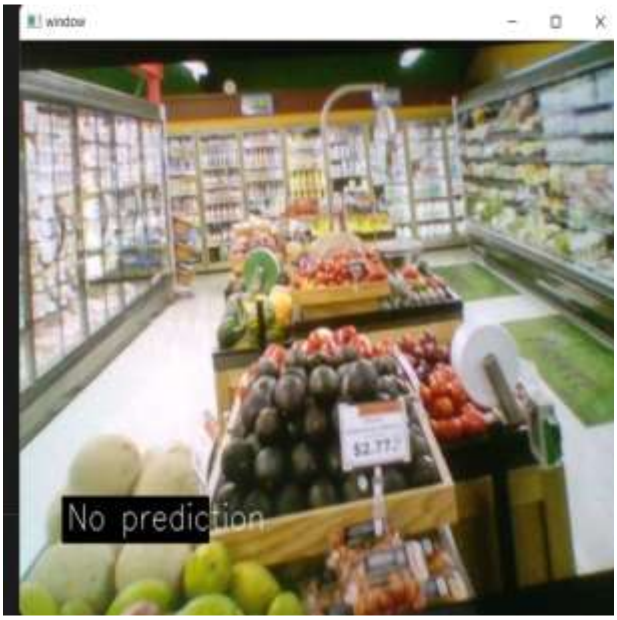


Fig. 7. Prediction of Supermarket Scene

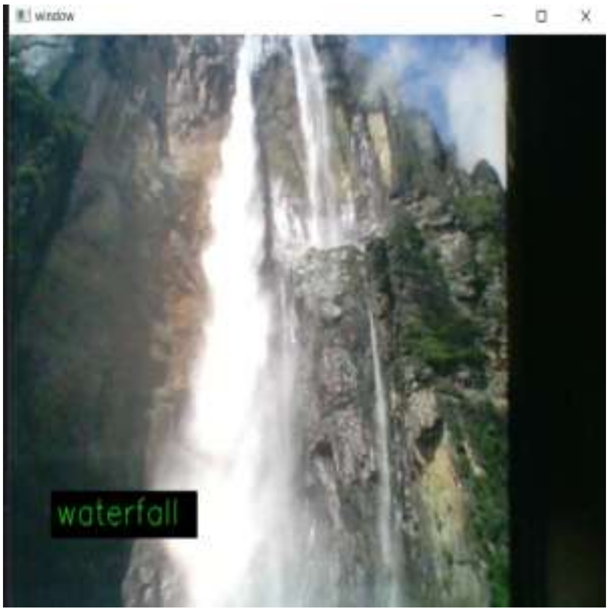
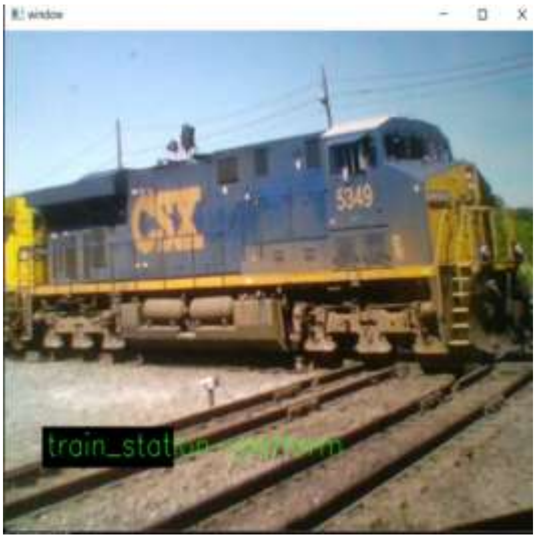
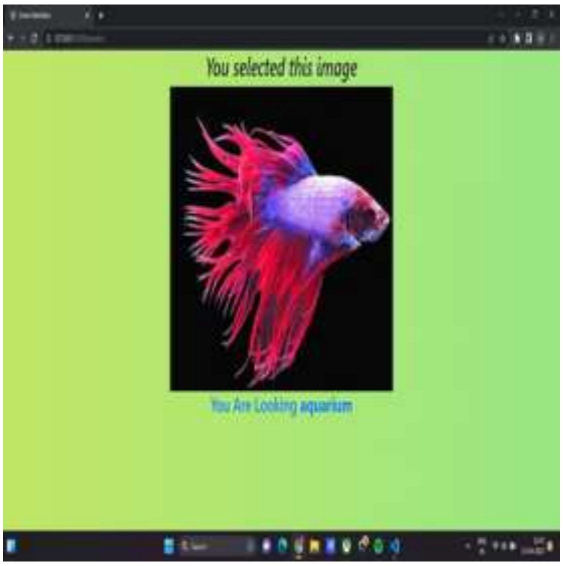


Fig. 9. Prediction of Waterfall Scene



train_station-platform

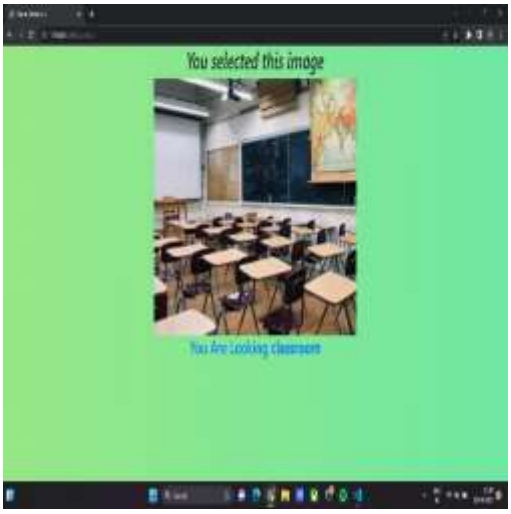
Fig. 8. Prediction of Train Scene



You selected this image

You Are Looking aquarium

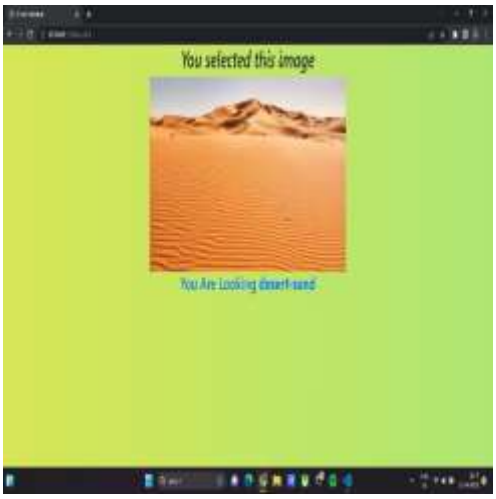
Fig. 10. Prediction of Aquarium Scene



You selected this image

You Are Looking classroom

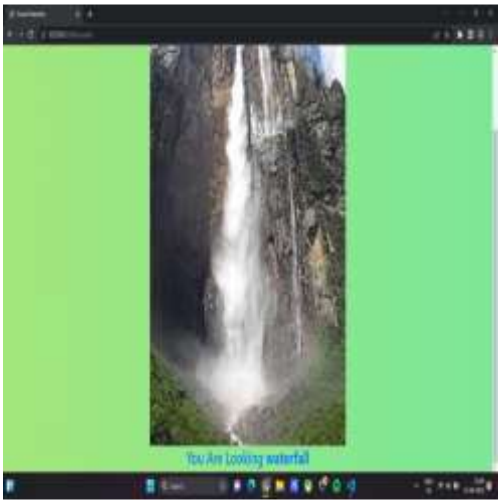
Fig. 11. Prediction of Classroom Scene



You selected this image

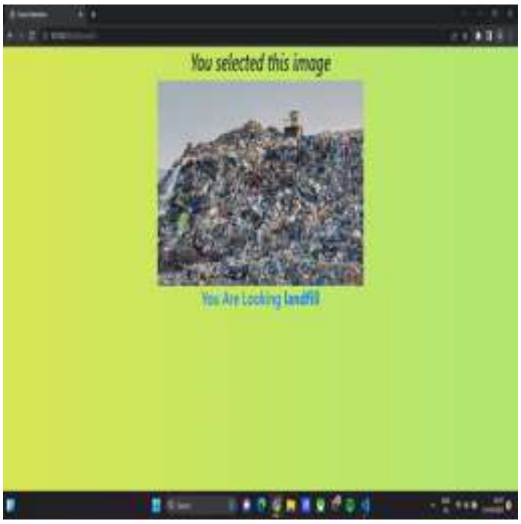
You Are Looking desert-sand

Fig. 13. Prediction of Desert Scene



You Are Looking waterfall

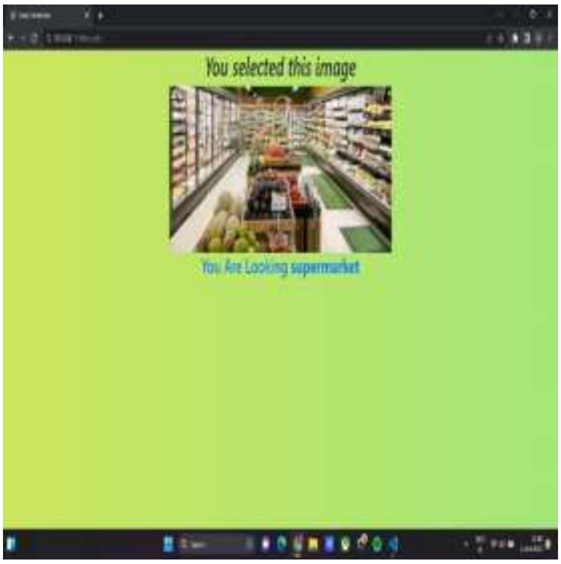
Fig. 12. Prediction of Waterfall Scene



You selected this image

You Are Looking landfill

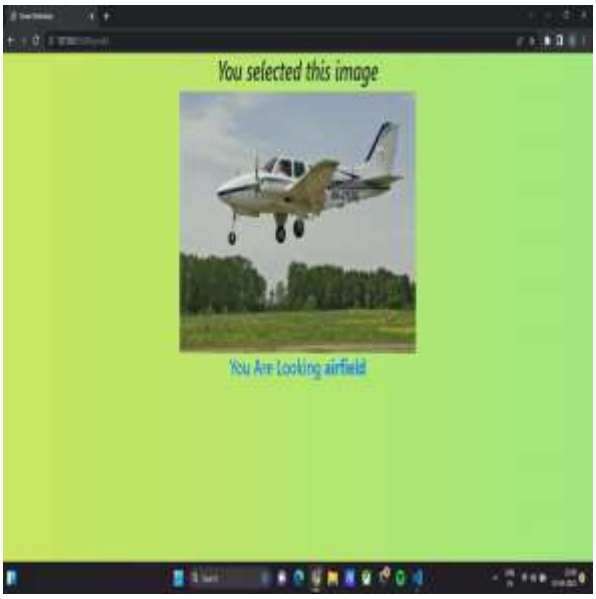
Fig. 14. Prediction of Landfill Scene



You selected this image

You Are Looking supermarket

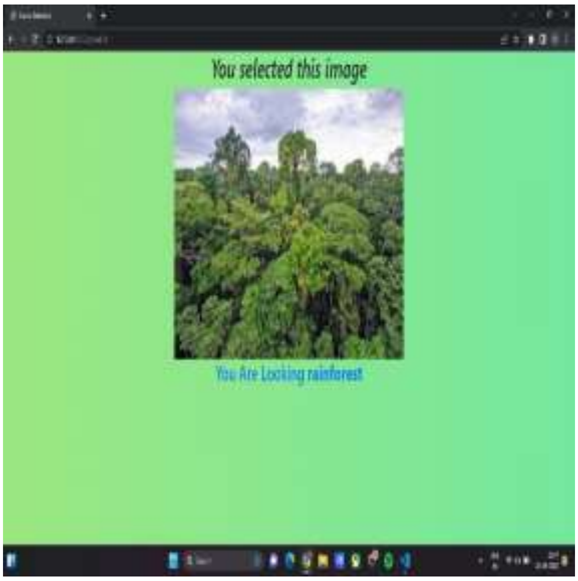
Fig. 15. Prediction of Supermarket Scene



You selected this image

You Are Looking airfield

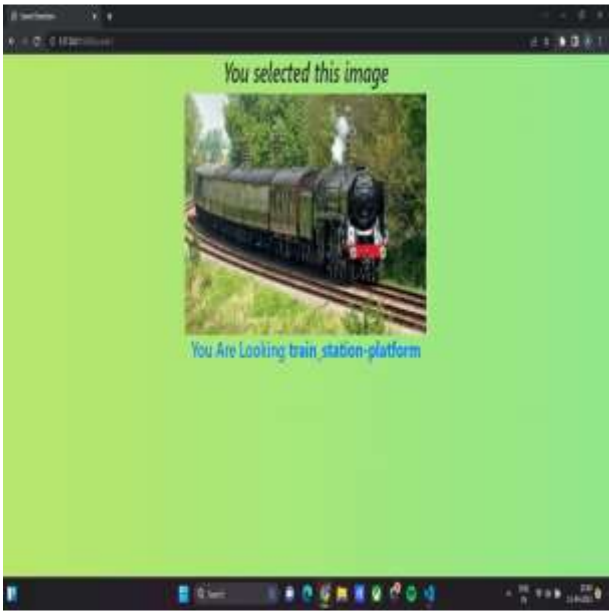
Fig. 16. Prediction of Airfield Scene



You selected this image

You Are Looking rainforest

Fig. 17. Prediction of Rainforest Scene



You selected this image

You Are Looking train_station-platform

Fig. 18. Prediction of Train Scene

VI. CONCLUSION AND FUTURE WORK

In this project we have used a CNN model for classification of images present in the dataset. This dataset is used for both training the model and testing the model with test images and for testing we have used images outside of the dataset also. The images used are RGB images. We got a accuracy of 75 Percentage, where we are able to classify the scene in ten different classes like Airfield, Aquarium, Supermarket, Desert, Waterfall, Bookstore, Landfill, Classroom, Train, Rainforest. The time required for computing or training this images is very high in comparison with normal other images and it requires good processing power. To increase the accuracy you can add more cnn layers and train the model in GPU which will add a better result and yield a better accuracy. In future enhancement we are going to use dataset like places365 or SUN397 which has over 400 different classes to classify. Also to classify and recognize the scenes in video.

REFERENCES

[1] Oh, Bongjin and Lee, Junhyeok,"A case study on scene recognition using an ensemble convolution neural network," 2018 20th International Conference on Advanced Communication Technology (ICACT).

[2] Zeng, Haitao and Song, Xinhang and Chen, Gongwei and Jiang, Shuqiang, "Learning scene attribute for scene recognition," IEEE Trans- actions on Multimedia,2019.

[3] Xie, Lin and Lee, Feifei and Liu, Li and Kotani, Koji and Chen, Qiu, ”Scene recognition: A comprehensive survey,” Pattern Recognition, 2020.

[4] Rawat, Waseem and Wang, Zenghui,”Deep convolutional neural net- works for image classification: A comprehensive review,” Neural computation, 2017.