Visual Object Recognition using Tensorflow

Pavan S Patil[1], Santhoshkumar V[2], Udaykumar V[3], Vishwas K K[4]
[1234] Electronics and Communication Department, KSIT College
Dr. B Sudarshan[5], Associate Professor
KSIT, Bengaluru, India

***Abstract*---Object Recognition is one of the most exciting areas in machine learning right now. Recognize objects like faces or cats are not so difficult, but recognizing arbitrary objects within a larger image set has been the difficult by artificial intelligence. The real surprise is that human brains recognize objects so well and effortlessly convert photons bouncing off objects into a spectacularly rich set of information about the objects around us. Machine learning still struggles with these simple tasks, but in the past few years, it's gotten much better.**

***Keywords*—Deep Neural Network (DNN), Convolutional Neural Network (CNN), Object recognition, Machine Learning, Deep Learning**

I.INTRODUCTION

Object Recognition technology has seen a quite good adoption rate in various and diverse industries. It helps in safe navigation of self-driving vehicles through traffic, spots violent behavior in a crowded place, it can used by sports teams for analyzing and build scouting reports, quality control of parts in manufacturing can be monitored and many other things. Deep learning is adopted so that brain neural network functions to perceive the data, such as image, sound can be simulated artificially. Deep learning by ImageNet has made an huge amount of progress toward object recognition by, collecting and processing those data sets. This paper involves analyzing various machine learning techniques for object recognition, so that different methodologies and algorithms used for object recognition can be achieved.

II.LITERATURE SURVEY

The paper [1] proposes deep learning approach in the recognition of objects in the historical building photographs of the town Trnava. It uses Deep learning architectures based on convolutional neural networks (CNN) for object recognition tasks. Cascade of convolution layers and activation functions are used to improve architecture. It is very important to setup of the number of layers and the number of neurons in each layer. TRNAVA LeNet 10 model was built and trained the purpose. This model is based on the dataset of 460 training images and 140 validation images which is the ratio of 3:1, images are of dimensions 28x28 pixels and image type used was color, image encoding was jpg. The model successfully recognized the right object in the photograph of historical building in Trnava. The proposed model gained 98.88% prediction accuracy.

The paper [2] proposes deep learning methods for facial expression recognition instead of hand-crafted features. Two kinds of deep networks such as deep neural network (DNN) and convolutional neural network (CNN) are used to solve recognition problems. The deep networks were developed using CUDA supported deep learning toolkits such as Caffe and CudaConvnet2 for high speed. Also, for implementing Haar-like face detection algorithm, they used OpenCV library. The images were cropped and resized to 64×64. Then, the 327 face images were divided to 10 groups, and then used one group for training and nine groups for test. The recognition results were good for 6 emotions, but the recognition rate of disgust label was poor. Because, the number of training images of disgust label in FER 2013 database was only 547. The DNN has the possibility of overfitting.

The paper [3] proposes considerable improvement in object detection and tagging using convolutional neural networks has given way to accurate yet complex methods, which can identify objects in real-time. However, the growth in the area of implementing the algorithms on low powered portable devices has been relatively slow, aims to converge the fields of computer vision and robotics, focusing on implementation of image description applications on an embedded system platform. The objects in the image are restricted to a fixed number, specific to data set used for training the model. According to Shaoqing Ren et al, the introduction of Region Proposal Network (RPN) allows sharing of whole image convolutional features with the network, thus, providing near cost-free region proposals. Wherein, region proposal technique is used to guide the algorithm in order to locate objects residing in an image. Secondly, execution of this method in our system allows the system to be computationally efficient and customized to run on low-powered machines.

The paper [4] proposes an object localization method to boost the performance of current object recognition techniques, utilizes the image edge information as a clue to determine the location of the objects. The Generic Edge Tokens (GETs) of the image are extracted based on the perceptual organization elements of human vision. These edge tokens are parsed according to the Best First Search algorithm to fine-tune the location of objects, where the objective function is the detection score returned by the Deep Convolutional Neural Network. Applying the BFS to the object localization and its search space, the search space is a set of edge

elements whose overlaps with the current candidate object is greater than zero. Testing the model in real time proved to be more efficient than the RCNN, also with scope for further development by improving the object localization by using a combination of the image edge, color and texture information, and the learned features of the image.

The paper [5] proposes a method to develop an interactive application in order to detect objects from videos, upon user input, it is also able to detect the particular object being shown at that instant on the screen. A sequential frame extraction method of videos and also deep learning approach of Convolutional Neural Networks along with Fully Connected Neural Networks is used for this task which gives an accuracy rate of 77%. Even when the object is somewhat distorted, translated, rotated or partially obstructed from view, it can be easily detected by humans, the task of computer vision is still quite challenging. Videos are made of frames synchronized with some playback audio, taking advantage of the fact that videos consist of frames, the analysis of the video can be made in much more detail by examining the object present in the frame images themselves, running the classifier and thereby get probabilities for different classes and hence classify the genre and also detect any object in the video. The operational accuracy of this model is improved by increasing the number of datasets and improving the hardware configuration so that the object classification can be done over a wider range of classes and in a faster way.

III. METHODOLOGY

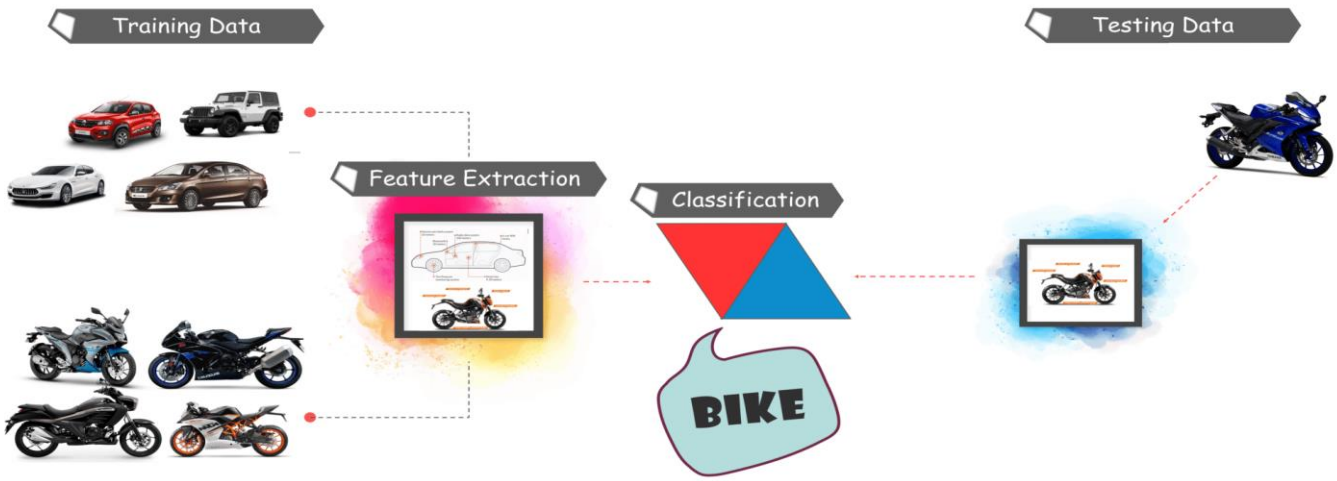


Fig 1.1 workflow

Every Object Detection Algorithm has a different way of working, but they all work on the same principle. They extract features from the input images at hands and use these features to determine the class of the image. Be it through MatLab, Open CV, Viola Jones or Deep Learning.

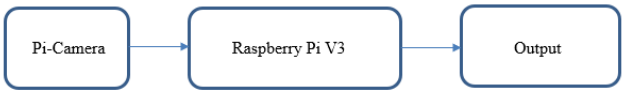


Fig 1.2 block diagram of object recognition

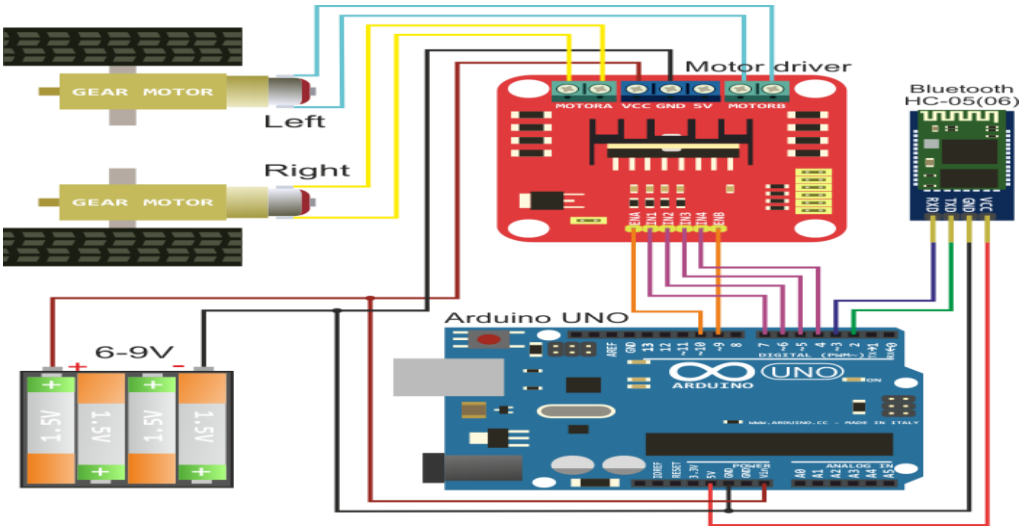


Fig 1.3 block diagram of 4WD motor wheel

Programming robot : First Step is to install TensorFlow, but TensorFlow actually comes with a makefile that lets us to build it right on the system. The steps take a few hours and have quite a few dependencies, but they work great.TensorFlow comes with a prebuilt model that performs object recognition.

Running : An image from the camera will output the top five guesses. If the camera is blocked and it gets a dark or blurry image it usually decides that it's looking at nematodes—clearly an artifact of the data it was trained on.

=>There are two approaches to performing object recognition using deep learning :

• Training a model from scratch: To train a deep network from scratch, you gather a very large labelled dataset and design a network architecture that will learn the features and build the model. The results can be impressive, but this approach requires a large amount of training data, and you need to set up the layers and weights in the CNN.

• Using a pretrained deep learning model: Most deep learning applications use the transfer learning approach, a process that involves fine-tuning a pretrained model. You start with an existing network, such as AlexNet or GoogLeNet, and feed in new data containing previously unknown classes. This method is less time-consuming and can provide a faster outcome because the model has already been trained on thousands or millions of images.

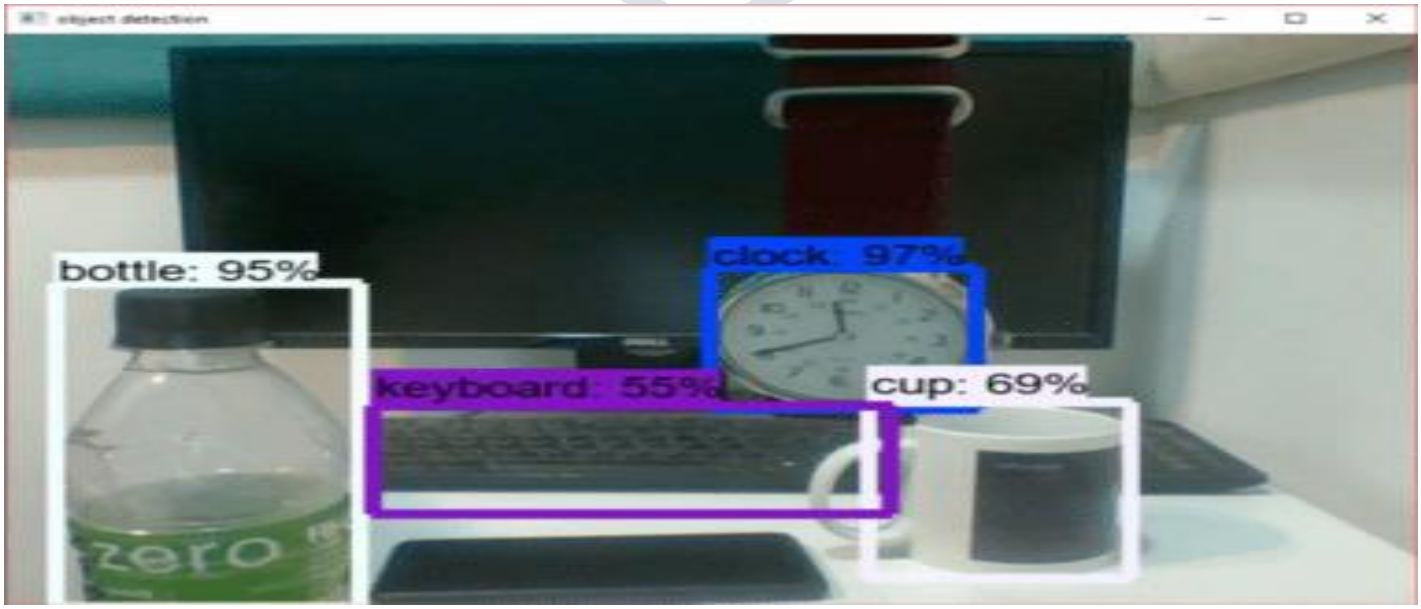
=>Other more basic approaches to object recognition may be sufficient depending on the application.

• Template matching – which uses a small image, or template, to find matching regions in a larger image.

• Image segmentation and blob analysis – which uses simple object properties, such as size, color, or shape. Typically, if an object can be recognized using a simple approach like image segmentation, it's best to start by using the simpler approach. This can provide a robust solution that does not require hundreds or thousands of training images or an overly complicated solution.

IV.RESULTS

Final output is displayed on Object detection window. Some of the output is shown below.



IV.CONCLUSION

Object recognition is one upcoming and exciting areas in machine learning. A well- known application of object detection is face detection, that is used in almost all the mobile cameras. But the major setback in all these papers are real time application and accuracy rate achieved, these can be overcome by Inception Model advanced architecture of ImageNet. These systems can be integrated with other tasks such as pose estimation where the first stage in the pipeline is to detect the object, and then the second stage will be to estimate pose in the detected region. It can be used for tracking objects and thus can be used in robotics and medical applications. Thus, this problem serves a multitude of applications. Deep learning models are the best considered for object detection because of training time, low latency, faster evaluation, etc.

V. REFERENCES

[1] Bezak, P. (2016, September). *Building recognition system based on deep learning*. In 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (pp. 1-5). IEEE.

[2] Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., & Ahn, C. (2015, January). *Development of deep learning-based facial expression recognition system*. In 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) (pp. 1-4). IEEE.

[3] Tenguria, R., Parkhedkar, S., Modak, N., Madan, R., & Tondwalkar, A. (2017, April). *Design framework for general purpose object recognition on a robotic platform*. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 2157-2160). IEEE.

[4] Etemad, E., & Gao, Q. (2017, September). *Object localization by optimizing convolutional neural network detection score using generic edge features*. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 675-679). IEEE.

[5] Mazumdar, M., Sarasvathi, V., & Kumar, A. (2017, August). *Object recognition in videos by sequential frame extraction using convolutional neural networks and fully connected neural networks*. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1485-1488). IEEE.

[6] Sujana, S. R., Abisheck, S. S., Ahmed, A. T., & Chandran, K. S. (2017, April). *Real time object identification using deep convolutional neural networks*. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 1801-1805). IEEE.

[7] Cheng, S. C. (2005). *Visual pattern matching in motion estimation for object-based very low bit-rate coding using moment-preserving edge detection*. IEEE transactions on multimedia, 7(2), 189-200.

[8] Alexeev, A., Matveev, Y., & Kukharev, G. (2018, October). *Using a Fully Connected Convolutional Network to Detect Objects in Images*. In 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 141-146). IEEE.

[9] Wong, S. C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I., & McDonnell, M. D. (2017). *Track everything: Limiting prior knowledge in online multi-object recognition*. IEEE Transactions on Image Processing, 26(10), 4669-4683.

[10] Yang, L., Wang, L., & Wu, S. (2018, April). *Real-time object recognition algorithm based on deep convolutional neural network*. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 331-335). IEEE.