International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

Traffic Sign Detection and Recognition using Deep Learning

Rahul Bhakad
Computer Engineering
Sinhgad Academy of Engineering
Savitribai Phule Pune University, Pune,
Maharashtra, India

Kunal Gaikwad
Computer Engineering
Sinhgad Academy of Engineering
Savitribai Phule Pune University, Pune,
Maharashtra, India

Shantanu Bawankule
Computer Engineering
Sinhgad Academy of Engineering
Savitribai Phule Pune University, Pune,
Maharashtra,India

***Abstract*—Traffic signs are important to ensure smooth traffic flow without bottle necks or mishaps. Traffic symbols are the pictorial representations having different necessary information required to be understood by the driver. Traffic signs in front of the vehicle are ignored by the drivers and this can lead to catastrophic accidents. To tackle this problem, we implemented a deep learning model based on YOLO's(You Only Look Once) latest version YOLOv5 in this project.YOLO is one of the fastest object detection algorithms for real-time detection.The goal is to find and classify traffic signs in real-world street settings.This paper presents an overview of the traffic sign board detection and recognition and implements a procedure to extract the road sign from a natural complex image, processes it and alerts the driver using voice command.**

Keywords:-YOLO, Convolution Neural Network, CNN, GTSDB, Deep Learning, Machine Learning, Object Detection, Text-To-Speech conversion, pyttsx3

INTRODUCTION

Every day, over 400 traffic accidents occur in India, according to official statistics. Road signs aid in the prevention of traffic accidents, assuring the safety of both drivers and pedestrians. Furthermore, traffic signals ensure that road users follow specified regulations, reducing the chances of traffic offenses. The usage of traffic signals also helps with route guidance. All road users, including automobiles and pedestrians, should prioritize road signals. For a variety of causes, such as focus issues, tiredness, and sleep deprivation, we overlook traffic signs. Poor vision, the impact of the outside world, and environmental circumstances are all factors that contribute to ignoring the signs.

As a result, several new machine learning approaches and algorithms have been developed to address all of these issues. TSD formerly relied heavily on classic object detection methods. TSD's pipeline often used hand-crafted features to extract region proposals, followed by the employment of classifiers to filter out the negatives. Deep learning approaches have risen in popularity in recent years, resulting in tremendous progress in target detection and identification tasks. Deep convolutional neural networks (CNNs) are used in the majority of image recognition and object detection studies to improve precision, speed, and accuracy. CNN can learn features from big datasets without preprocessing, avoiding the need for hand-crafted features and absorbing more generalized features.

CNN has been employed in recent advances in object identification algorithms such as SSD, Fast R-CNN, Faster R-CNN, R-FCN, and YOLO. We employed a single-shot detection network called "You Look Only Once" (YOLO), which has low propagation delay and good detection performance. Many current neural networks are accurate but do not operate in real time, necessitating the use of a large number of GPUs for training[1]. YOLO solves these issues by constructing a CNN that runs in real time on a standard GPU and only requires one conventional GPU for training[1].

Using state of the art technology, YOLOv5 ('You only look once'), which is an object detection algorithm that divides images into a grid system. Each cell in the grid is responsible for detecting objects within itself. YOLOv5 is one of the most famous object detection algorithms due to its speed and accuracy. This technology will examine images acquired by a car's front-facing camera in real time and assist the driver by raising concerns to him or her through audio output or vehicle navigation display.

RELATED WORK

There are two primary components in traffic sign recognition systems: classification and recognition. After a traffic sign has been spotted, classification means determining the type of traffic sign, and recognition means locating the traffic sign in a succession of images. TSD approaches are classified into two types. The first is based on classic object detection algorithms, which are based on traffic sign features such as color-based and shape-based traffic sign features. The other is learning-based technologies, such as machine learning and deep learning techniques, which can self-learn numerous characteristics.

Low-level feature extraction techniques are used in traditional ways to recognise or detect signs, which require primary colors and shape features. However, these methods

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

were confined to a few basic types of traffic signs. The algorithms for detecting the location of traffic sign instances in photos in these methods were either time-consuming or focused solely on the categorization of pre-cropped traffic sign instances. Machine learning-based detection approaches were later introduced.

Feature extraction techniques have been used extensively in traditional detection approaches. Both classification and detection rely on characteristics such as color and form.Images are usually remodeled into HSV to beat RGB color house limitations for varied light-weight condition.[2] presents a color chance model that computes maps supported Ohta space. the overall shapes of the traffic signs embody circle, triangle, parallelogram or the other plane figure. Contours extracted from edges along side the options of color and shape are used. In traditional detection ways, histogram of orientated (HOG) feature with SVM classifier are state of art techniques. [3] employs the HOG options along side Associate in Nursing SVM classifier that gave a decent performance on the German Traffic Sign Detection Benchmark (GTSDB)competition hosted by IJCNN in 2013.Even though manually picked features have achieved improved precision for traffic signs.Traditional detection methods are relatively particular and lack robustness towards changing scenes and the complications that come with them.

After 2013, CNNs were mostly employed for TSD and TSR research. Modern detectors have two parts: a backbone and a head. The backbone is pre-trained on ImageNet, while the head predicts object classes and bounding boxes. VGG, ResNet, ResNeXt, or DenseNet could be the backbone for detectors that run on GPU systems. SqueezeNet, MobileNet, or ShuffleNet could be the backbone for detectors that run on the CPU platform. There are two types of object detectors in the head: one-stage and two-stage object detectors.[4] The R-CNN family, which includes quick R-CNN, Faster R-CNN, R-FCN, and Libra R-CNN, is the most common two-stage object detector and YOLOv5 as a one-stage object detector.

Although R-CNNs can produce the needed results, their computation efficiency is poor, and the process takes a long time and requires a lot of resources. Because of the unified network structures, single-stage approaches are substantially faster.

YOLO is an abbreviation meaning You Only Look Once. We're using YOLOv5, which is the most advanced object recognition algorithm currently available. It's a brand-new convolutional neural network (CNN) that accurately recognises objects in real time. YOLO is extraordinarily rapid, with a mean average precision (mAP) that is more than double that of conventional real-time systems[5]. This method processes the entire image with a single neural network, then divides it into parts and predicts bounding boxes and probabilities for each component. The predicted probability is used to weight these bounding boxes.The approach "looks once" at the image In the sense that it produces predictions after only one forward propagation through the neural network.

As a result, we can conclude that one-stage approaches are effective for obtaining faster results with good precision. Because traffic sign detection must be done in real time, the YOLOv5 would be a better choice.

OUTCOME OF LITERATURE REVIEW

As previously stated, various models for traffic sign detection were investigated, as well as their benefits and drawbacks. In comparison to other models, the YOLO model was found to be the most fit for the application due to its speed and lowest test time delay.

PROPOSED SOLUTION

To develop a YOLO based deep CNN model for Traffic Sign Detection and Classification, trained on the German Traffic Sign Dataset. The goal is to get a model which can identify and group traffic signs continuously and coordinating it with pyttsx3 python text-to-speech library to give a voice alert to the driver or the traveler at whatever point the traffic sign is detected.

PREREQUISITE

A. DATASET

Our custom dataset contains total 8946 images of traffic signs,our dataset is divided into two parts,one is for training and another one is for testing.There are total 7800 images for training and 1146 images for testing.Images are annotated for identifying location and the class of traffic signs in the image.The dataset contains natural traffic signs pictures, shot on different kinds of streets (roadway, rustic, metropolitan) during the daytime, around sunset, and different weather patterns which makes traffic signs suffer from difference in orientation, light conditions, or occlusions.There are total 36 classes of traffic signs.Hence our model can identify different types of traffic signs on the road and thus making our model very reliable.

B. PYTTSX3(PYTHON TEXT-TO-SPEECH)

PYTTSX3 is a Python-based text-to-speech conversion library.It is supported by many operating systems and operates offline,unlike other libraries, and is compatible with Python 2 and 3.It is a simple tool that turns the text you type into speech.PYTTSX 3 can be installed with pip package manager.After installed PYTTSX3 will load the driver according to the operating system.It includes sapi5 on Windows,nsss on MacOs and espeak on linux.The PYTTSX3 module offers two voices,the first of which is female and the second of which is male.It also allows to change the rate of speech and volume according to the needs.

INTRO TO YOLOv5

The YOLO models are end-to-end deep learning models that are preferred for their speed and accuracy in detecting objects. The structure of a YOLO network is similar to that of a traditional CNN, with numerous convolutional and max-pooling layers leading to two fully connected layers. YOLO is a novel method in that it uses a single neural network to look at the full image only once, giving it a number of advantages over classifier-based systems. It uses characteristics from the full image to predict each bounding box.

YOLO predicts bounding boxes and probability of expected item classes using a single neural network. The photos are only pass through the network once.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

S × S grids are used to split the image to be input. With C conditional class probabilities, each grid cell predicts k bounding boxes and confidence scores for these bounding boxes. A quintuple (x, y, w, h, cf d) characterize each of the bounding boxes. The (x, y) coordinates provide the center offset of the bounding box The (x, y) coordinates provide the center offset of the bounding box compared to the bounds The width and height of the predicted object with relation to the entire image are represented by the variables w and h.The confidence cf d is defined as Pr(Object) ∗ IOU(truth/pred).

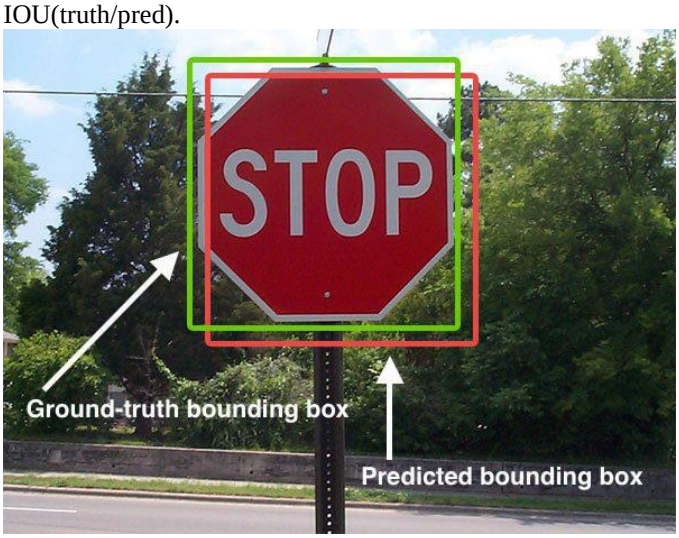

Figure 1: Intersection Over Union

When a grid cell contains a component of a ground truth box, Pr(Object) is 1, otherwise it is 0. The intersection over union (IOU) between the predicted box and its matching ground truth box is a metric that ranges from 0 to 1, with 0 indicating no overlap and 1 indicating that the predicted box is identical to the ground truth. For all bounding boxes in a given region, there is only one set of class scores C. As a result, the YOLO network produces an output for each image that is a vector of S × S × (5B + C) numbers. The cf d and class probabilities are combined to determine which grids will be chosen for prediction.We calculate the class-specific confidence score of individual bounding boxes with the help of these predictions, and then select the bounding boxes with high confidence scores in each grid cell to produce global predictions of a traffic sign in the image.

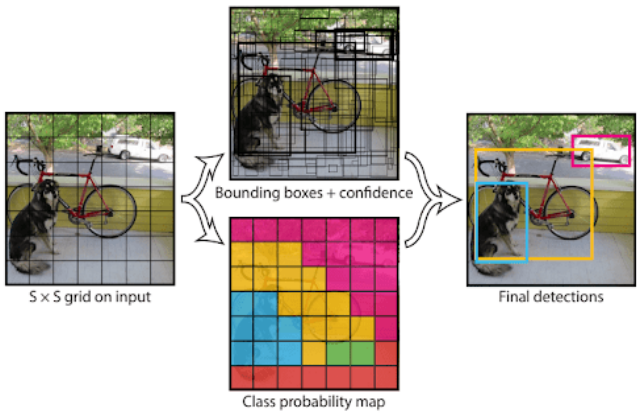

Figure 2: YOLO object detection process

YOLOv5 ARCHITECTURE

YOLOv5 is a single-stage object detector, like any other single-stage object detector it contains three key components. Model Backbone,Model Neck and Model Head.

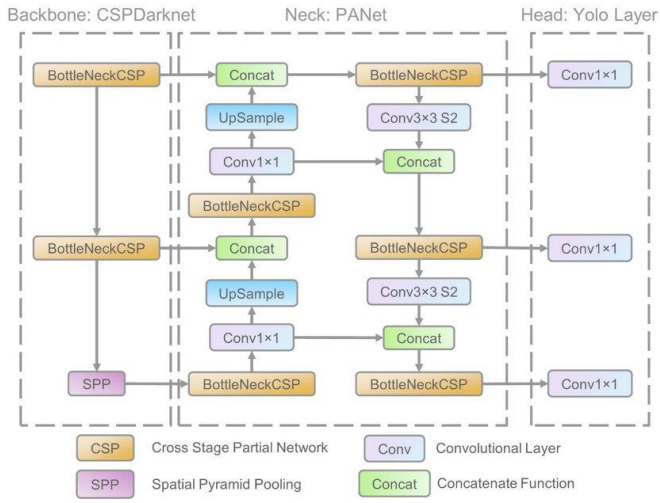

Figure 3: YOLOv5 Architecture

1) Backbone:The basic purpose of Model Backbone is to extract key features from an input image. The CSP (Cross Stage Partial Networks) backbone is utilized in YOLO v5 to extract rich in useful features from an input image.
2) Neck:The main purpose of Model Neck is to generate feature pyramids. Feature pyramids aid in the generalization of models on object scaling. It aids in the identification of the same object in various sizes and scales. Feature pyramids are quite helpful in assisting models in performing effectively on unknown data. Other models, such as FPN, BiFPN, PANet, and others, employ various sorts of feature pyramid approaches. PANet is utilised as the neck in YOLO v5 to obtain feature pyramids.
3) Head:The model Head is primarily utilized for the last stage of detection. It creates final output vectors with class probabilities, objectness scores, and bounding boxes after applying anchor boxes to features.

Activation Function:The Leaky ReLU and Sigmoid activation function were chosen by the YOLO v5 creators. In YOLO v5, the middle/hidden layers use the Leaky ReLU activation function, whereas the final detection layer uses the sigmoid activation function.

Optimization Function:For optimization function in YOLO v5, we've alternatives.SGD and Adam.In YOLO v5, the default optimization feature for training is SGD.but, you can change it to Adam by using using the "—— adam" command-line argument.

TRAINING DISCUSSION

We used Google's Colab Notebook to train our model. Colab is a browser-based tool that allows users to write and run python code. It's great for machine learning, data analysis,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 03, March-2022

and education. More precisely, Colab is a hosted Jupyter notebook service that requires no installation and provides free access to computational resources and GPUs. Colab has a GPU with a capacity of 12GB. The types of GPUs accessible for customers to use in Colab change over time. This is frequently required for Colab to be able to provide free access to certain resources. Nvidia K80, T4, and P100 GPUs are available from Colab.

MODEL

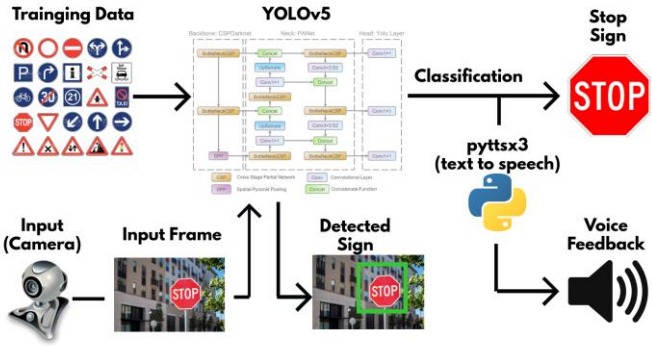

Figure 4: YOLOv5 Trained Model Implementation

1) Training:Model is first trained with the traffic signs dataset
2) Input data: After training we will be using a camera as input device to feed video frames to the trained model.
3) Prediction: The trained TSDR model can then process the input frames to detect and recognize the traffic signs and create bounding boxes around the traffic signs with the label of the detected traffic sign and
4) Voice Feedback: The class prediction of the traffic signs detected in every frame will be a string e.g. “stop”. By using pyttsx3 we can then convert these strings to speech,and through the speaker we can produce the voice feedback of the detected traffic sign.

CONCLUSION

We explored the topic of detecting and recognising a large number of traffic-sign categories in this presentation with the goal of automating traffic-sign recognition and informing the driver via text or audio output. We proposed using a technique called YOLOv5 algorithm and how it is employed in traffic sign detection. When compared to other object detection approaches such as Fast R-CNN, Faster R-CNN and other R-CNN algorithms, this technique delivers faster detection results. The system uses a deep network to learn a huge number of categories while also detecting them efficiently and quickly. During the picture acquisition step, the photos will be captured with a camera and the detection will be done using the YOLOv5. When a traffic sign is detected, the system provides a voice alert. This model is best suited when requiring precise and safe navigation.

REFERENCES

[1] Bochkovsky, Alexey, Chien-Yao Wang and H. Liao. “YOLOv4: Optimal Speed and Accuracy of Object Detection.” ArXiv abs/2004.10934 (2020): n. pag.
[2] G. Wang, G. Ren, and T. Quan, “A traffic sign detection method with high accuracy and efficiency,” in Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering (ICCSEE), Hangzhou, China, 2013, pp. 22–23.
[3] G. Wang, G. Ren, Z. Wu, Y. Zhao, and L. Jiang, “A robust, coarse-to-fine traffic sign detection method,” in In Proceedings of the 2013 International Joint Conference on Neural Networks, Dallas, TX, USA, 2013, pp. 754–758
[4] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA,,2014, pp. 580–587
[5] Redmon, J., Farhadi, “A.: YOLO9000: better, faster, stronger”. In: IEEE CVPR, pp. 7263–7271 (2017)
[6] Adrian Rosebrock, Intersection over Union (IoU) for object detection on November 7, 2016.https://pyimagesearch.com/
[7] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi.”You Only Look Once: Unified, Real-Time Object Detection” arXiv:1506.02640v5
[8] Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A Forest Fire Detection System Based on Ensemble Learning. Forests 2021, 12, 217.