

# Contrasting YOLO Models for the Formulation of Diaphanous and Robust Object Detection System

Abhishek Singh

Department of Computer Science and  
Engineering,  
School of Engineering & Technology,  
Sharda University  
Greater Noida, India

Ronit Attrey

Department of Computer Science and  
Engineering,  
School of Engineering & Technology,  
Sharda University  
Greater Noida, India

Anuj Kumar

Department of Computer Science and  
Engineering,  
School of Engineering & Technology,  
Sharda University  
Greater Noida, India

**Abstract** – The Aid for Blind People has been one of the Prioritised Outlook of the Governments and the Tech Sector. With an Exponential Increase in the Aggregate of Automobiles on Road due to Ever Escalating Population, Commuting on Road Has Become a Strenuous Task for Normal Human Being as Well. Considering these Factors, We Tried to Explicate a System which would Help the Blind People in Commuting on Roads by Some Extent. In Order to Formulate Something which would be Easier to be Integrated into the Regular Well-Being of the Visually Impaired, In this Paper, We've Endeavoured the State-of-the-Art YOLO Models i.e., YOLOv5 and YOLOv8 to Scrutinize the Impact of the Complexity of these Models on a Particular Dataset. With this Inclusion, we'll be able to assuage a System which would be Diaphanous and Authentic for our Real-time Use Case. Road Signs and Signals being one of the Most Dynamic Aspect of Commute, We Induced it as our Custom Dataset, for Understanding the Behaviour of the Models. With this Research, We Understood the Differences between a Light-Weight and an Accurate System, Possessing an Ease of Applicability. Both the Models Exhibited High Level Performances on our Dataset and Gave us a Resolution which is Potent and Reliable for the Development of an Overall System for Prevention of the Visually Impaired, from any Misfortune on Roads.

**Keywords** — Deep Learning, You Only Look Once (YOLO), Object Detection, YOLOv5, YOLOv8, Road Sign and Signal.

## I. INTRODUCTION

In the Domain of Artificial Intelligence, the Inclusion of any Modular Approach into a Real-time System is only possible if that Particular System is Light-Weight, Accurate, Less Complex and Reliable. As per World Health Organization (WHO), Around 2.2 Billion People are Visually Impaired across the Globe [1]. Moreover, as per the Research Carried out in University of California, around 20% Blinds in the World, Face an Unintentional Mobility Accidents Which Causes Unwanted Deaths [2]. As a Result, the Creation of System for the Blinds, to Ease their Commute stands to be the Demand of Time. With the Initiation of the Process of Incorporating a System with Hardware and Software, the Real Challenge Lies

in Gauging a Software System which could be Easily Imbided into a Hardware and could be Used by the Visually Impaired, in their Day-to-Day Workflow. The Basic Object Detection Flow with the Help of Deep Learning Methodologies Elucidates Itself into 05 Major Points [4]:

- The Acquisition of Data.
- The Preparation of Data.
- Detection of Region of Interest (ROI).
- Training of the Model.
- Gauging the Performance of a Model through a Metric.

Figure.01, Demonstrates the Modularity of Object Detection System.

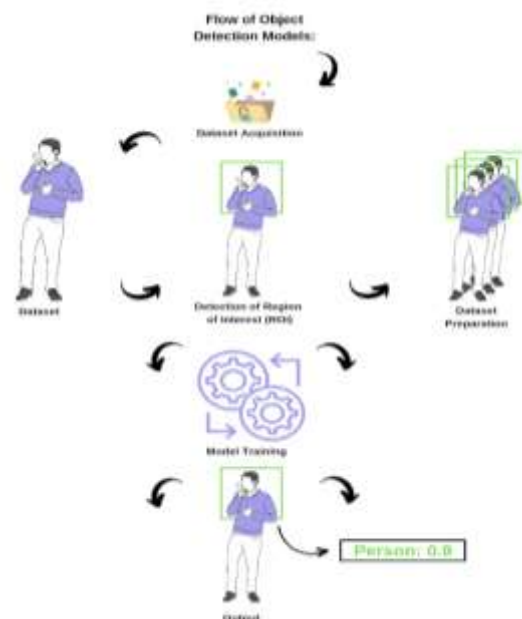


Fig.1. Fundamental Flow of an Object Detection System.

After Understanding the Object Detection Flow, We Got to know that the Modular Phenomenon Plays a Major Role in Formulating a Real-time System. Thus, We wanted a Light-Weight and Robust Algorithm for Object Detection. As a Result, we Selected State-of-the-Art YOLOv5 Model [6], which has Proven its Efficacy in Varied Object Detection Utilities and YOLOv8 Model [20], which is the Latest and the More Complex Model of the You Only Look Once (YOLO) Family. The Ultimate Reason to Select these Models over other Modular Approaches was the Nature of YOLO Models, i.e., They're Light-Weight, Highly Accurate and Reliable for Real-time Applicability. After Selecting the Contrasting Models, As We wanted to Ease the Road Commute of the Visually Impaired, We Collated Road Signs and Signal Dataset for Testing the Model's Performance. The Dataset Consist of Varied Road Signs on Normal Roads and the Traffic Signals making it a Dataset of 17 Classes. Now, After the Inclusion of Well-Defined Modular Phenomena and Well-Explicated Dataset as per the Use-case, We Tested our Overall Approach. The Results which we got were Astonishing and Gave us a Clear Ideation of the Differences Between the Less Complex and More Complex Modular Approach.

The Exhibited Modus Operandi, Continues with the Scrutiny of Prevalent Technological Ideations in Section-II. Furthermore, Section-III, Elucidates the Detailed Analysis of Methodological Proposition and Its Actuation. Moreover, the Analysis of the Experimental Results is Carried out in Section-IV and the Modular Outlook is Concluded with Potent Insight and Relevant Future Scope in Section-V.

## II. RELATED WORK

The Modular Approaches for Object Detection (Encompassing Road Safety) Has Been Widely Carried Out. Majorly, We Tried to Focus our Review on Models and the Objects that are related to Vehicular Prospect, i.e., Vehicle Detection, Road Sign and Signal Detection, License Plate Recognition, to Understand the Modus Operandi in each of these Phenomenon's. Thus, A Generous Comparison was Carried out between YOLOv2, YOLOv3 and Single Shot MultiBox Detectors (SSDs) over Road Sign Plates as a Dataset, Gauging an Average Mean Absolute Precision (mAP) of 90%. But There was a Scope of Improvement over Dataset (Which were Less in Number) and Accuracy which is Less for any Real-time Application. [3]. The Formulation of a System for Detection of Occluded Signs was Amalgamated with Convolutional Neural Network (CNN) Model, Procuring a Maximum Precision of 96.34%. The Model's Performance could've been Better if a Greater Number of Occluded Images would have been Added for Training [4].

A Similar Modular Outlook with the Induction of Convolutional Neural Network (CNN) over Belgium Road Signs Data was also proposed, Exhibiting an Accuracy of 83.7% Overall, which is less when Compared to Already

Existing Methodologies [5]. For aiding the Advanced Driver Assistance Systems (ADAS), a Modular Approach with Belgium and German Road Signs Data Trained over Convolutional Neural Network (CNN), to Achieve an Accuracy of Above 90% on an Average and Could be Increased if a Greater Number of Data would've been Induced for Training [8]. A Modus Operandi Encompassing Video Sequencing Technique with Inculcation of Deep Neural Networks for Traffic Sign Detection was Carried out and the Author's got an Average Accuracy of 95% which is Good Enough, when compared to other Approaches, but the Model Becomes really Complex and Heavy due to which Real-time Applicability Deteriorates Drastically [9]. For Vehicle Detection Modus Operandi, YOLOv3 was Used, procuring an Accuracy of 98% with 10,000 Instances of Data. The Increased Number of Data, Escalated the Efficacy of the Model and Gave Better Results [10]. Moreover, the License Plates of Bangladesh was Taken as a Data and was Trained over Convolutional Neural Network (CNN) Network with Addition of 02 Unique CNN Layers for Better Accuracy, and got a False Rate of 0.025 which was Calculated Manually [11]. Furthermore, for the Detection of Myanmar License Plates, OpenALPR System was Explicated over Basic Dataset of Myanmar License Plates and Expelled an Average Accuracy of 90%. The Confidence Score of the Modus Operandi would've been Improved with the Induction of More License Plate Data into the System [12].

A Rigorous Comparative Analysis Between Faster-RCNN, YOLOv3 and Single Shot MultiBox Detector (SSD) has also been carried out for the Purpose of Detection of Vehicles. The Efficacy in the form of Mean Absolute Precision (mAP), which each Modular Approaches Exhibited were 0.75 of RCNN, 0.52 of SSD and 0.56 of YOLOv3. The Mean Absolute Precision (mAP) of any of these Approaches were not as Expected and thus, the Author Tuned the RCNN Model for More Potent Mean Absolute Precision (mAP) Value and Exceeded it from 0.75 to 0.82 [13]. The Usage of Raspberry Pi along with the Utilization of OpenCV for Vehicle Detection was Incorporated, gauging an Accuracy of 95%, which stands to be the Average of all the Predictions. This Approach is Cost-Effective, but it Possesses the Constraint of Practicality When it comes to the Perspective of Inclusion of the Proposed System into a Real-time Use-Case [16].

As a Result, The Explications over Varied Aspect of Data and the Integration of these Data over a Well-Defined Network was Successfully Done. With these Explorations, we had a Clear Idea of the Models and the Data we need to Induce for Our Modus Operandi. We got to Know the Impact of Constraints of Data, Complexity of the Model, Etc., on the Method's Performance. In our Approach, We Included all the Crucial Short-Comings found in the Review and tried to Avoid those for Procuring Better Efficacy with Keeping in Mind the End Goal of Real-time Actuation and Aiding the Visually Impaired.

Thus, the Data, the Model and the Accuracy Metric was Chosen for our Modus Operandi after the Detailed Explication.

### III. PROPOSED METHODOLOGY & ITS IMPLEMENTATION

The Methodological Outlook was Created with Deep Implorations, Taking into Consideration the Basics of Object Detection which is Demonstrated Visually in Figure.01. Moreover, Figure.02, Depicts the Proposed Flow of our Modus Operandi.

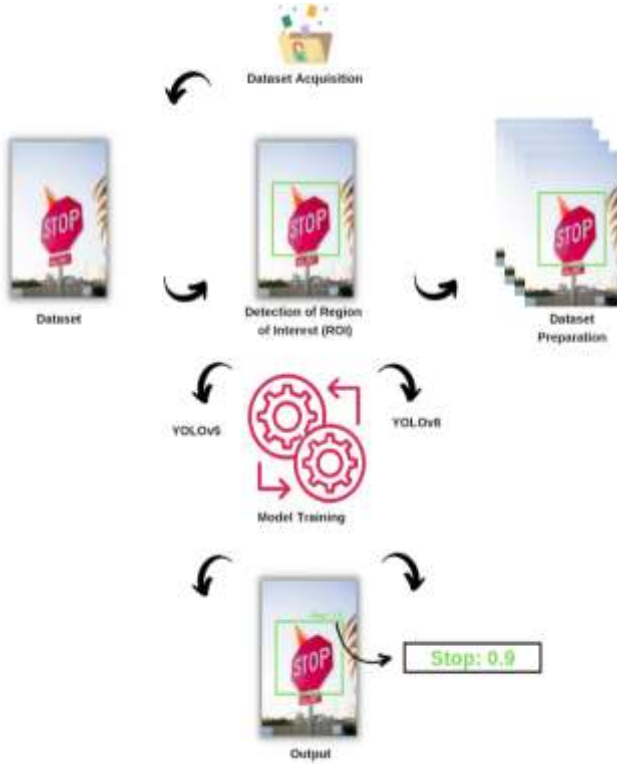


Fig.2. The Visual Interpretation of the Proposed Flow.

Let's Explicate each Step in Detail,

**A. The Data Acquisition and Preparation:** The Dataset which we Have Utilised in our Modus Operandi is Custom Made. We took Various Road Signs and Signals Instances from the Internet and Amalgamated them to form a Strength of 2093 Images with 17 Classes. The Median Image Ratio is 640x640 as we're Dealing with YOLO Models and Average Image Size is 0.41 Mega Pixel (MP). Figure.03, Illustrates Some of the Images of the Data. The 17 Classes of Data are Well-Balanced and Possess Nearly Equal Number of Images for Each Class, providing us a Justified Result if Trained on any Network. Each Image is Labelled and a Bounding Box (BB) [11] is Created over each Images Region of Interest (ROI) [14]. The Same is Depicted in Figure.03.

With all the Prominent Information of the Dataset, We Tried to Formulate each Class of the Data in such a Way that it Possess

Not Only the Clear Images but also the Occluded Images to make the Training of the Model Robust and Reliable [12].



Fig.3. Illustration of Road Signs and Signal Dataset.

The Representation of Balanced Dataset which is an Important Aspect of Modular Training is Depicted in Figure.04. Apart from "green\_light" Data Class and somewhat "red\_light" Data Class, all the Classes are Well-Balanced in the Formulated Dataset.



Fig.4. Demonstration of Balanced Dataset.

**B. The Training of Object Detection Modular Networks:** After the Initial Step of Data Acquisition and Data Preparation, now it's Time to Train the Selected Models over the Same. As we Selected YOLOv5 and YOLOv8 Models from the YOLO Family for Training, Let's Understand the Modular Aspect of Each of These Models and also the Contrasting Factor Between Them.

**YOLOv5 Model:** You Only Look Once (YOLO) v5 is Utilized in this Research, as it Exhibits State-of-the-Art Results in the Domain of Object Detection. The Architecture is Based upon the Concept of Deep Learning, and among all the YOLO Models in the YOLO Family, YOLOv5 [18] is Simplistic and Reliable. The Requirement of Computational Strength for the Model Training is also Extremely Less, which Makes it a Model with Less Complexity. Although the Utilization of Computational Power is Less in YOLOv5, the Results it Gives is Comparable with Other Complex Networks. With all this Checkpoints for our Modus Operandi, YOLOv5 Performs Extremely Faster when Compared to other Networks. YOLOv4 [14] Architecture is Potently Induced in YOLOv5 with the

Integration of Cross Stage Partial CSPDarknet as an Encoder. With this Encoder and the Path Aggregation Network (PANet), an Entire Architecture of the YOLOv5 Network is Formulated. The Activation Function in the YOLOv5 is Replaced from Leaky ReLU and Hard Swish Activations (Utilized in YOLOv4) [15] to Sigmoid Linear Units (SiLU) Activation Function. Figure.

05, Demonstrates the Block Architecture of YOLO's Network Formulation.

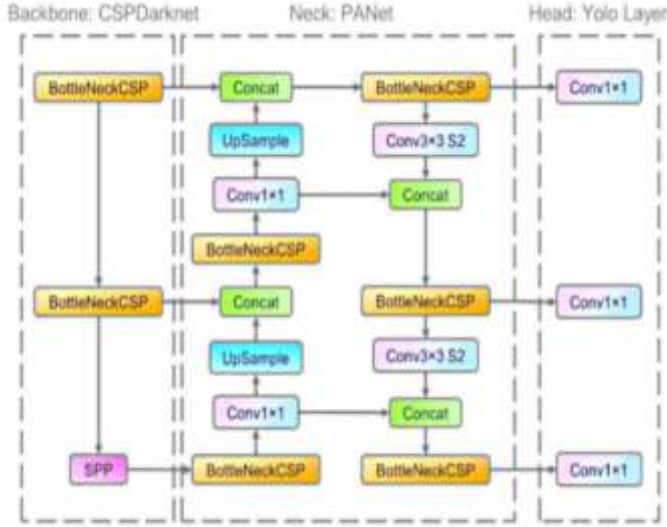


Fig.5. The Block Architecture of YOLO Network [17].

By Understanding the Potency of YOLOv5 Network, we Trained it Over our Dataset for Over 100 Epochs, Considering the Optimal Batch Size of 32 and Learning Rate as 0.01.

*You Only Look Once (YOLO) v8*: YOLOv8 [20] is the Latest Outcome of the YOLO Family with Certain Structural Refinements. It's Built over the Framework of YOLOv5. The Center of the Object is Predicted Directly, rather than the Offset through a Known Anchor Box, as YOLOv8 is an Anchor Free Model. With this Enhancement, there is Deterioration in the Prediction of Boxes, which thereby Escalates the Non-Maximum Suppressions (NMS) [14]. Moreover, a 6x6 Conv at the Stem is Changed to 3x3. The Initial Conv's Kernel Size of the Bottleneck has been Changed from 1x1 to 3x3 and the Rest is Same as in YOLOv5 [17]. With this Integration, YOLOv8 has Shown an Inclination towards the ResNet System Flow.

Furthermore, Mosaic Augmentation has also been Introduced in YOLOv8 Model which augments the Instances at the Time of Training Itself. With Each Increasing Epoch, the Model Visualizes a Slightly Altered Images, which Does the Job of Data Augmentation and which thereby Improves the Result. The Process is Carried out by Amalgamation of 04 Images Intact, Enabling the Model to assimilate the Objects into New Locations, Where the Images are Partially Occluded and

Opposing to Varied Pixels in the Surrounding of an Image. Figure.06, Depicts the YOLOv8 System Architecture in Detail. As the YOLOv8, is Extremely New and Improved Architecture, we Induced it for Our Training and did for Over 100 Epochs, with an Optimal Batch Size of 32 and the Learning Rate of 0.01.

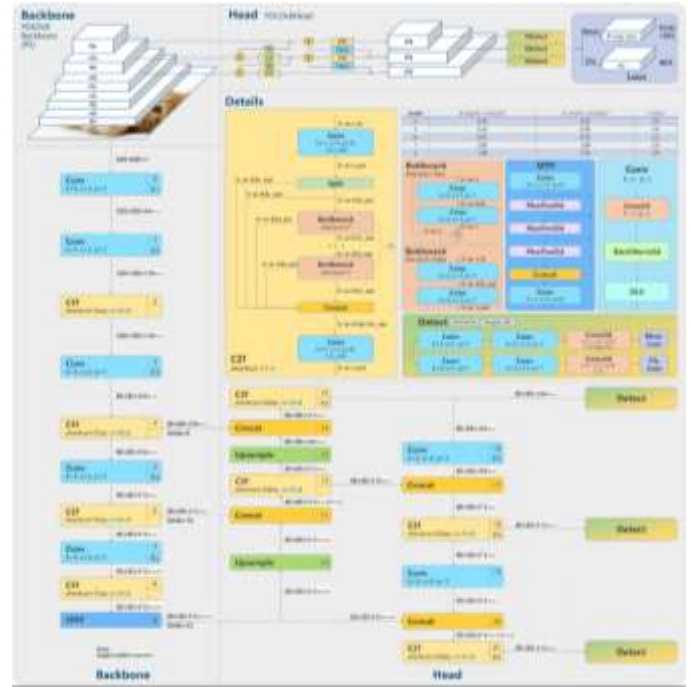


Fig.6. The Block Architecture of YOLOv8 Network [20].

*YOLOv5 vs YOLOv8*: In the Deep Learning Domain, there are Various Object Detection Modus Operandi , but YOLOv5 and YOLOv8, are the Best Models created by Ultralytics. YOLOv5 is Well-Known for its Simplicity, Speed and Efficacy, Whereas YOLOv8 being the Latest Addition into the YOLO Family has Improved itself with the Induction of Performance Spike and have become more Flexible. But when Contrasting Both, YOLOv5 is Light-weight in Nature, then YOLOv8. In Section-IV, we'll Understand the Impact of these Models on our Dataset in Detail.

*C. Mean Absolute Precision (mAP)*: The mAP [15] is one of the Most Famous Metric utilized for Gauging the Accuracy of an Object Detection Model which is Calculated with the Inclusion of Following Sub-Metrics i.e., Confusion Matrix, Intersection over Union (IoU) [21], Recall and Precision. The Value of mAP Lies Between 0 to 1. The Formula for mAP is Demonstrated Below, where 'AP' is the Average Precision which is Calculated for Each Class Initially and then it's Averaged over the Overall Number of Classes in a Dataset.

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad \dots (1)$$



Equation.01 [21] Demonstrates the mAP which Induces Substitution Between Recall and Precision and Incorporates False Positives (FPs) and False Negatives (FNs) Both. This Makes mAP a Suitable Metric for Detection Algorithms.

With this Flow, we Constituted our Pipeline for Object Detection, In Order to Juxtapose the Best Models and thereby Procuring a Model with Maximum Efficacy and Efficiency.

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

After the Implementation We Effectuated, The Results which we Got were Highly Insightful and we got the Desired Analysis, which we were Expecting.

**A. YOLOv5 Analysis:** On Training YOLOv5 Model over our Road Sign and Signal Dataset, the mAP which we Obtained were, a) mAP50: 96% and, b) mAP50-95: 80%. The Mean Absolute Precision (mAP) which we got is Really Promising, when Compared to Other Existing Modular Prospects. Figure.07, Depicts the Graphical Illustration for YOLOv5 Performance.

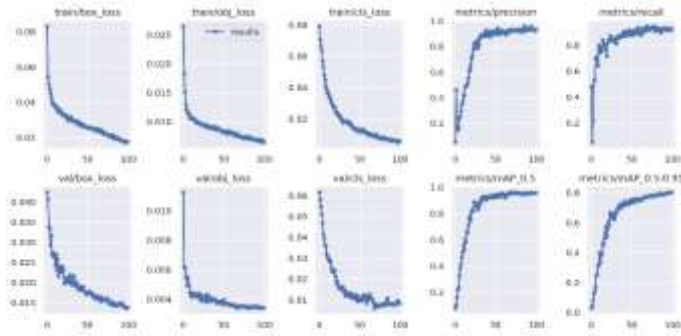


Fig.7. Detailed Illustration of Performance of YOLOv5 Model over Our Dataset.

As our Model is of 17 Classes, the Accuracy for Each of the Class has also been Deduced using YOLOv5. Figure.08, Represents the Confusion Matrix for YOLOv5.

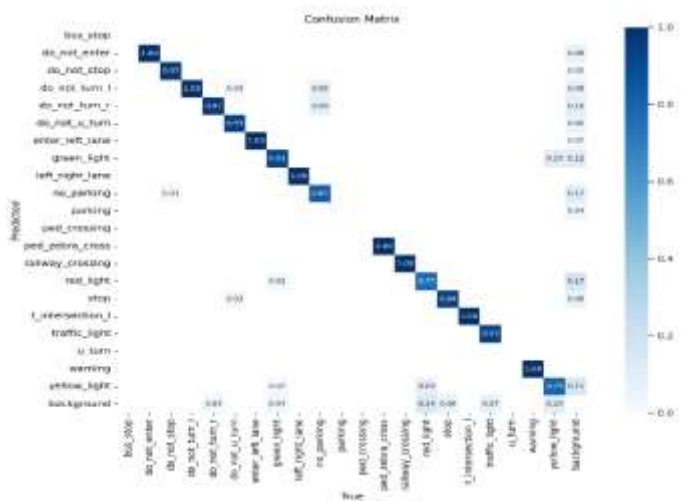


Fig.8. Confusion Matrix for YOLOv5.

**B. YOLOv8 Analysis:** On Training YOLOv8 Model over our Road Sign and Signal Dataset, the mAP which we Obtained were, a) mAP50: 94.5% and, b) mAP50-95: 79.5%. The Mean Absolute Precision (mAP) which we got is Really Promising, when Compared to Other Existing Modular Prospects. Figure.08, Depicts the Graphical Illustration for YOLOv5 Performance.

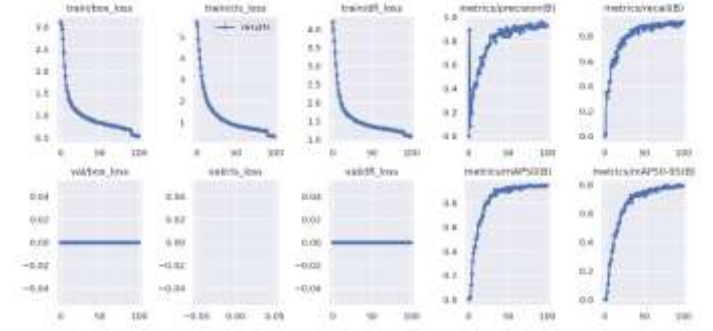


Fig.7. Detailed Illustration of Performance of YOLOv8 Model over Our Dataset.

As our Model is of 17 Classes, the Accuracy for Each of the Class has also been Deduced using YOLOv8. Figure.09, Represents the Confusion Matrix for YOLOv8.

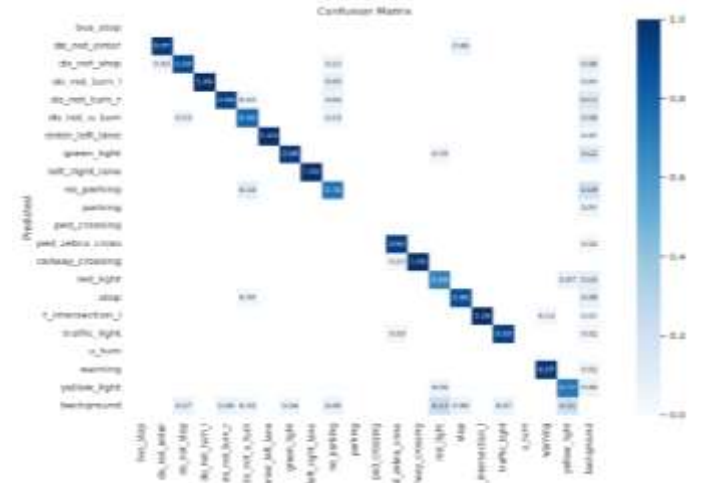


Fig.9. Confusion Matrix for YOLOv8.

With the Results which we Got from YOLOv5 and YOLOv8 and Comparing their Accuracies [11], with already Existing Methodologies, We were assured of Choosing these Models for our Purpose. As our Dataset is of 17-Classes, due to which, Often the Performance of the Model Deteriorates, Our Approach and Tuning has Resulted into a Potent Outlook for Object Detection for Visually Impaired. Table-I, Depicts the Mean Absolute Precision for Each Class of Our Data, Trained on YOLO Networks i.e., YOLOv5 and YOLOv8.

Thus, Our Experimental Analysis was Successful and Gave us a Clear Insight of the Model which we would be Using as a Software which is Covered in Section-V.

TABLE I. ANALOGIZING THE MEAN ABSOLUTE PRECISION OF YOLOV5 AND YOLOV8 OVER 17 CLASSES.

S.No.	Class Names	YOLOv5 mAP (in %)	YOLOv8 mAP (in %)
1.	do_not_enter	99.5	98.9
2.	do_not_stop	96.9	95.2
3.	do_not_turn_l	98.9	99.5
4.	do_not_turn_r	98.2	95.9
5.	do_not_u_turn	98.4	94.8
6.	enter_left_lane	96.7	98.4
7.	green_light	93.2	95.9
8.	left_right_lane	99.5	99.5
9.	no_parking	97.6	88.4
10.	ped_zebra_crossing	99.5	99.4
11.	railway_crossing	99.5	99.5
12.	red_light	83.4	78.6
13.	stop	96.5	93.9
14.	t_intersection_l	99.5	99.5
15.	traffic_light	96.6	94.6
16.	warning	99.5	99.5
17.	yelllow_light	76.8	69.1

## V. CONCLUSION AND FUTURE SCOPE

As our End Goal is to Create a Real-time System to aid the Visually Impaired, we needed a Model which would be Light-Weight, Fast and Reliable. After Analysing the Overall and Class-Wise Results of Both the YOLO Models, YOLOv5 Performed Better with Our Dataset Even being Less Complex, when Compared to YOLOv8. The Performance of Both the Models are Top-Notch, but for our Modus Operandi, YOLOv5 Suits the Best and also Gave us the Insight that the Complexity of the Models isn't Directly Proportional to the Accuracy. The Whole Flow, Depends Upon the Dataset and the Tuning which is Arbitrarily Done. Furthermore, With the Integration of YOLOv5 along with the Text to Speech Amalgamation, a Wearable Can be Introduced for the Visually Impaired to Be Safe on Roads and thereby Ease their Commute as Well. There's always a Scope of Improvement and More Dataset and More Tuning are some of the Most Probable Areas where you can Increase the Model's Efficacy Even More.

## REFERENCES

[1] <https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment>  
[2] <https://users.soe.ucsc.edu/>  
[3] Y. -L. Kuo and S. -H. Lin, "Applications of Deep Learning to Road Sign Detection in DVR Images," 2019 IEEE International Symposium on

Measurement and Control in Robotics (ISMCR), Houston, TX, USA, 2019, pp. A2-1-1-A2-1-6, doi: 10.1109/ISMCR47492.2019.8955719.  
[4] J. Guo, X. Cheng, Q. Chen and Q. Yang, "Detection of Occluded Road Signs on Autonomous Driving Vehicles," 2019 IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019, pp. 856-861, doi: 10.1109/ICME.2019.00152.  
[5] V. Swaminathan, S. Arora, R. Bansal and R. Rajalakshmi, "Autonomous Driving System with Road Sign Recognition using Convolutional Neural Networks," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-4, doi: 10.1109/ICCIDS.2019.8862152.  
[6] J. Yu, P. Ge and D. Zhang, "Research on Detection and Recognition Algorithm of Road Traffic Signs," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 1996-2001, doi: 10.1109/CCDC.2019.8833426.  
[7] Y. Sun, P. Ge and D. Liu, "Traffic Sign Detection and Recognition Based on Convolutional Neural Network," 2019 Chinese Automation Congress (CAC), Hangzhou, China, 2019, pp. 2851-2854, doi: 10.1109/CAC48633.2019.8997240.  
[8] A. Vennelakanti, S. Shreya, R. Rajendran, D. Sarkar, D. Muddegowda and P. Hanagal, "Traffic Sign Detection and Recognition using a CNN Ensemble," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2019, pp. 1-4, doi: 10.1109/ICCE.2019.8662019.  
[9] I. Belkin, S. Tkachenko and D. Yudin, "Traffic Sign Recognition on Video Sequence using Deep Neural Networks and Matching Algorithm," 2019 International Conference on Artificial Intelligence: Applications and Innovations (IC-AIAI), Belgrade, Serbia, 2019, pp. 35-354, doi: 10.1109/IC-AIAI48757.2019.00013.  
[10] D. Chowdhury, S. Mandal, D. Das, S. Banerjee, S. Shome and D. Choudhary, "An Adaptive Technique for Computer Vision Based Vehicles License Plate Detection System," 2019 International Conference on OptoElectronics and Applied Optics (Optronix), 2019, pp. 1-6, doi: 10.1109/OPTRONIX.2019.8862406.  
[11] N. Saif et al., "Automatic License Plate Recognition System for Bangla License Plates using Convolutional Neural Network," TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON), 2019, pp. 925-930, doi: 10.1109/TENCON.2019.8929280.  
[12] K. P. P. Aung, K. H. Nwe and A. Yoshitaka, "Automatic License Plate Detection System for Myanmar Vehicle License Plates," 2019 International Conference on Advanced Information Technologies (ICAIT), 2019, pp. 132-136, doi: 10.1109/AITC.2019.8921286.  
[13] C. C. Paglinawan, A. N. Yumang, L. C. M. Andrada, E. C. Garcia and J. M. F. Hernandez, "Optimization of Vehicle Speed Calculation on Raspberry Pi Using Sparse Random Projection," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), 2018, pp. 1-6, doi: 10.1109/HNICEM.2018.8666325.  
[14] V. S. Sindhu, "Vehicle Identification from Traffic Video Surveillance Using YOLOv4," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1768-1775, doi: 10.1109/ICICCS51141.2021.9432144.  
[15] F. H. Shubho, F. Iftekhar, E. Hossain and S. Siddique, "Real-time traffic monitoring and traffic offense detection using YOLOv4 and OpenCV DNN," TENCON 2021 - 2021 IEEE Region 10 Conference (TENCON), 2021, pp. 46-51, doi: 10.1109/TENCON54134.2021.9707406.  
[16] A. P. Kulkarni and V. P. Baligar, "Real Time Vehicle Detection, Tracking and Counting Using Raspberry-Pi," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 603-607, doi: 10.1109/ICIMIA48430.2020.9074944  
[17] [https://www.researchgate.net/figure/The-network-architecture-of-Yolov5-It-consists-of-three-parts-1-Backbone-CSPDarknet\\_fig1\\_349299852](https://www.researchgate.net/figure/The-network-architecture-of-Yolov5-It-consists-of-three-parts-1-Backbone-CSPDarknet_fig1_349299852)  
[18] C. R. Rashmi and C. P. Shantala, "Vehicle Density Analysis and Classification using YOLOv3 for Smart Cities," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2020, pp. 980-986, doi: 10.1109/ICECA49313.2020.9297561.  
[19] A. P. Kulkarni and V. P. Baligar, "Real Time Vehicle Detection, Tracking and Counting Using Raspberry-Pi," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2020, pp. 603-607, doi: 10.1109/ICIMIA48430.2020.9074944.  
[20] <https://docs.ultralytics.com/>  
[21] <https://www.v7labs.com/blog/mean-average-precision>